SAND2022-13427
Printed September 2022

Sandia National Laboratories

Harmonized Automatic Relay Mitigation of Nefarious Intentional Events (HARMONIE) – Special Protection Scheme (SPS)

Sandia National Laboratories: Shamina Hossain-McKenzie, Nicholas Jacobs, Adam Summers, Bryan Kolaczkowski, Chris Goes, Ray Fasano

Texas A&M University: Zeyu Mao, Leen Al Homoud, Kate Davis, Thomas Overbye

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

The harmonized automatic relay mitigation of nefarious intentional events (HARMONIE) special protection scheme (SPS) was developed to provide adaptive, cyber-physical response to unpredictable disturbances in the electric grid. The HARMONIE-SPS methodology includes a machine learning classification framework that analyzes real-time cyber-physical data and determines if the system is in normal conditions, cyber disturbance, physical disturbance, or cyber-physical disturbance. This classification then informs response, if needed and/or suitable, and included cyber-physical corrective actions. Beyond standard power system mitigations, a few novel approaches were developed that included a consensus algorithm-based relay voting scheme, an automated power system triggering condition and corrective action pairing algorithm, and a cyber traffic routing optimization algorithm. Both the classification and response techniques were tested within a newly integrated emulation environment composed of a real-time digital simulator (RTDS) and SCEPTRE™. This report details the HARMONIE-SPS methodology, highlighting both the classification and response techniques, and the subsequent testing results from the emulation environment.

ACKNOWLEDGEMENTS

The HARMONIE-SPS team would like to express their deepest gratitude to all the subject matter experts, past and present project contributors, and mentors for making this research meaningful and successful. Specifically, we would like to thank former team member Daniel Calzada for his foundational work in developing the classification framework design and initial implementation. We would also like to thank former team member Hanyue Li on her novel work on developing the autoRAS approach. We would also like to thank the Public Service Company of New Mexico for their invaluable time for discussing remedial action schemes and providing feedback on HARMONIE-SPS, especially the consensus algorithm-based relay voting scheme design. Collaboration with the ADROC team has been extremely beneficial for advancing both R&D efforts. Lastly, we would like to thank Matt Reno, Jason Stamp, Komal Shetye, and various consultants to the project in helping inform the initial HARMONIE-SPS idea, design, and implementation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

This page left blank

EXECUTIVE SUMMARY

With the advent and integration of novel smart grid technologies that broaden the cyber attack surface, the rise of unpredictable disturbances such as electromagnetic pulses (EMPs), and the looming presence of extreme weather events, a next-generation special protection scheme (SPS) with the following attributes is needed:

1. A SPS that can adapt to unpredictable events (without predefined conditions) and effectively respond to limit/eliminate the disruption quickly
2. A SPS that is cyber-physical in analyzing collected data and taking response actions; it is no longer sufficient for a SPS to process only physical power system data and solely take physical-side actions; cyber-side actions are necessary to eliminate malicious compromise
3. A SPS that extends the use of protective relays from fault isolation to also adaptively learning system conditions, preventing cyber attack propagation, and taking proactive actions to prevent compromise within the relay set itself

To meet the needs of future SPSs, we propose a defensive, wide-area SPS that learns system conditions, mitigates cyber-physical consequences, and preserves grid operation under diverse predictable and unpredictable disturbances. This harmonized automatic relay mitigation of nefarious intentional events (HARMONIE)-SPS will meet the needs stated above by processing both cyber and physical data from both relays and out-of-band (OOB) measurements, learning actual system conditions to adapt to both predictable and unpredictable disturbances, and performing proactive response actions to prevent further cascading impact. Furthermore, the HARMONIE-SPS will leverage the distributed sets of protective relays, within different zones, to derive classification of the system conditions and respond to disturbances. With this increased situational awareness and proactive control response approach, the HARMONIE-SPS can greatly improve the resilience of the electric grid against cyber-physical disturbances, whether it is intentionally malicious or inadvertent.

In this report, we detail our HARMONIE-SPS approach, including the machine learning framework, cyber-physical testbed development, and consensus algorithm-based relay voting (CARV) scheme. We also explore the different types and combinations of cyber-physical corrective actions that can be deployed, including automated pairing of triggering conditions and physical (power system) corrective actions, cyber network routing optimization, etc. Lastly, we detail collaboration with the ADROC: An Emulation Experimentation Platform for Advancing Resilience of Control Systems LDRD project for assessing CARV within the HARMONIE-SPS emulation environment.

For the experiments conducted to evaluate HARMONIE-SPS methodology, two different environments were used. One was the Texas A&M University RESLab cyber-physical environment that was used to model different disturbances in a WSCC 9-bus system use case and provide training and testing data for the initial machine learning classification framework. Second was the newly integrated RTDS and SCEPTRE™ cyber-physical emulation environment where two different use-cases were implemented: WSCC 9-bus system and IEEE 39-bus system. Different disturbances were implemented in this emulation environment to assess both the HARMONIE-SPS classification and response mechanisms. Results for these different disturbance scenarios within the two different environments are provided in this report and highlight HARMONIE-SPS's ability to process and classify cyber-physical electric grid data and utilize it to inform more comprehensive cyber-physical response that improves the system conditions and overall grid resilience.

ACRONYMS AND TERMS

Acronym/Term	Definition
ADROC	ADvancing Resilience Of Control systems
ADMM	Alternating Direction Method of Multipliers
AUC	Area Under the Curve
BDD	Bad Data Detection
BFT	Byzantine Fault Tolerance
CA	Consensus Algorithm
CARV	Consensus Algorithm-Based Relay Voting
CORE	Common Open Research Emulator
CPS	Cyber Physical System
DER	Distributed Energy Resource
D-FACTS	Distributed Flexible AC Transmission System
DNP3	Distributed Network Protocol 3
DoS	Denial of Service
EMS	Energy Management System
EMT	Electromagnetic Transient Simulation
ESA	Easy SimAuto
FCI	False Command Injection
FDI	False Data Injection
FGSM	Fast Gradient Sign Method
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
HARMONIE	Harmonized Automatic Relay Mitigation of Nefarious Events
HELICS	Hierarchical Engine for Large-scale Infrastructure Co-Simulation
HIL	Hardware In-The Loop
I	Current
ICS	Industrial Control System
IDS	Intrusion Detection System
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
LDRD	Laboratory Directed Research and Development

Acronym/Term	Definition
LSTM	Long Short Term Memory
MCC	Matthew's Correlation Coefficient
MHA	Multi-Head Attention
MITM	Man In The Middle
ML	Machine Learning
MVA	Mega Volt-Amps
MW	Mega Watts
NERC	North American Electric Reliability Corporation
OPC	Open Platform Communications
P	Real Power
PBFT	Practical BFT
PDC	Phasor Data Concentrator
PDE	Partial Differential Equation
PGD	Projected Gradient Descent
PMU	Phasor Measurement Unit
PNNL	Pacific Northwest National Laboratory
PV	Photovoltaic
PW	PowerWorld
Q	Reactive Power
RAS	Remedial Action Scheme
RBFT	Robust BFT
RNN	Recurrent Neural Network
ROC	Receiver Operator Curve
RTDS	Real Time Digital Simulator
RTT	Round Trip Time
RTU	Remote Terminal Unit
SCADA	Supervisory Control And Data Acquisition
SDN	Software Defined Networking
SDP	Shortest Distance Path
SE	State Estimation
SEL	Schweitzer Engineering Laboratories
SIMD	Single Instruction Multiple Data
SLG	Single-line-to-ground
SNL	Sandia National Laboratories
SPS	Special Protection Scheme

Acronym/Term	Definition
SVM	Support Vector Machine
TAMU	Texas A&M University
TD	Time Delay
TRAST	Transformative Remedial Action Scheme Tool
TS	Transient Simulation
WECC	Western Electricity Coordinating Council
WSCC	Western System Coordinating Council
V	Voltage
2oo3	Two-out-of-three
2oo4	Two-out-of-four

1. INTRODUCTION

Protection schemes are vital to the continuous, reliable operation of the electric grid. There exist a variety of protection schemes that coordinate protective relays during power system faults and seek to isolate and clear the faults quickly and efficiently to prevent any sustained damage and cascading impact. Extending the focus from isolating and clearing faults, special protection schemes (SPSs) protect the grid by detecting predefined abnormal conditions and deploying predefined corrective actions in a playbook manner. It is important to note that SPSs and remedial action schemes (RASs) terminology is often used interchangeably [1].

SPSs prioritize reliability and seek to maintain stability, acceptable voltages, and loading limits during disturbances, essentially operating within the respond and recover functions of National Institute of Standards and Technology's Cybersecurity Framework [2]. Unlike typical protection schemes, SPSs can take actions beyond the isolation of a fault and include changes to demand, generation, and system configuration.

However, it no longer suffices for SPSs to focus solely on predefined disturbances and reliability. Resilience and unpredictable disturbances such as electromagnetic pulses (EMPs), extreme weather, and malicious events threatening national security must be considered. Hurricane Maria in Puerto Rico revealed the fragility of grid physical infrastructure, which suffered severe damage and continues to require significant restoration effort. It showed how quickly cascading failures can destabilize even undamaged equipment and the dependency and interconnectedness of critical infrastructure [3].

Cyber attacks targeting grid operations are increasing in frequency and intensity, as exemplified by the calamitous 2015 and 2016 cyber attacks to the Ukrainian grid [4]. The recent Colonial pipeline cyber attack highlighted the cascading impact of compromises in critical infrastructure; hackers were able to use a compromised password to access the company's network and conduct a ransomware attack that caused the pipeline to shut down the system for the first time in their 57-year history. Cyber-physical systems such as the electric grid can have cascading impacts across the cyber and physical domains, a characteristic common in all critical infrastructure systems [5].

Furthermore, with the increasing penetration of distributed energy resources (DER) such as solar photovoltaic (PV) systems and wind farms, new technologies are being integrated and connected to the bulk power system. These grid-edge devices, with novel communication and automation functionalities, are also becoming targets to cyber attacks and can cause detrimental impact propagation as DER penetration increases [6].

With the advent and integration of novel smart grid technologies that broaden the cyber attack surface, the rise of unpredictable disturbances such as EMPs, and the looming presence of extreme weather events, a next-generation SPS with the following attributes is needed:

4. A SPS that can adapt to unpredictable events (without predefined conditions) and effectively respond to limit/eliminate the disruption quickly
5. A SPS that is cyber-physical in analyzing collected data and taking response actions; it is no longer sufficient for a SPS to process only physical power system data and solely take physical-side actions; cyber-side actions are necessary to eliminate malicious compromise
6. A SPS that extends the use of protective relays from fault isolation to also adaptively learning system conditions, preventing cyber attack propagation, and taking proactive actions to prevent compromise within the relay set itself

To meet the needs of future SPSs, we propose a defensive, wide-area SPS that learns system conditions, mitigates cyber-physical consequences, and preserves grid operation under diverse predictable and unpredictable disturbances. This harmonized automatic relay mitigation of nefarious intentional events (HARMONIE)-SPS will meet the needs stated above by processing both cyber and physical data from both relays and out-of-band (OOB) measurements, learning actual system conditions to adapt to both predictable and unpredictable disturbances, and performing proactive response actions to prevent further cascading impact.

Furthermore, the HARMONIE-SPS will leverage the distributed sets of protective relays, within different zones, to derive classification of the system conditions and respond to disturbances. With this increased situational awareness and proactive control response approach, the HARMONIE-SPS can greatly improve the resilience of the electric grid against cyber-physical disturbances, whether it is intentionally malicious or inadvertent.

In this report, we detail our HARMONIE-SPS approach, including the machine learning framework, cyber-physical testbed development, and consensus algorithm-based relay voting (CARV) scheme. We also explore the different types and combinations of cyber-physical corrective actions that can be deployed, including automated pairing of triggering conditions and physical (power system) corrective actions, cyber network routing optimization, etc. Lastly, we detail collaboration with the ADROC: An Emulation Experimentation Platform for Advancing Resilience of Control Systems LDRD project for assessing CARV within the HARMONIE-SPS emulation environment.

2. NEED FOR ADAPTIVE, CYBER-PHYSICAL SPS

In this section, discussion on traditional implementations of SPS/RAS is provided as well as how unpredictable disturbances challenge their effective operation. Additionally, existing automated response efforts and their foci are described.

2.1. Special Protection Schemes

According to the Western Electricity Coordination Council (WECC), there are four common elements for the design of SPS/RAS: arming criteria, initiating conditions, actions taken, and time requirements [7]. The arming criteria are critical system conditions for which a step-wise SPS should be ready to take action when required. The initial conditions are the contingencies that have been known to cause violations of reliability and stability standards, which will initiate the SPS corrective action if the scheme is armed. The initial conditions can be event-based, parameter-based, response-based, or the combination of the above.

Event-based schemes directly detect outages and/or fault events and initiate actions to mitigate the event consequence fully or partially. Parameter-based schemes measure variables for which a significant change confirms the occurrence of a critical event. Response-based schemes monitor system response during events and disturbances and incorporate a closed-loop process to react to actual system conditions [8]. The work of [9] finds that most SPS in the WECC system initiate upon changes to system topology, with very few being triggered by system condition changes.

2.2. Disturbances

Traditional SPSs, as described in the last section, are designed to combat predictable disturbances such as generator/line outages or faults that cause violations in system parameters such as voltage current, reactive power, etc. These disturbances, or contingencies, are extensively studied in the planning phase (e.g., N-1 contingency analysis) with simulation-based tools to understand what violations (e.g., triggering conditions) result for different combinations of contingencies and what mitigations can address them. The playbook design of these traditional SPSs works very well for predictable disturbances, especially with accurate models of the power system for simulation-based studies.

Disturbances such as cyber attacks and extreme weather have unpredictable trajectories that are difficult to predict and characterize in terms of triggering conditions as described in the previous subsection. Certain known cyber attacks can have signatures and behavior-based indicators; however, the attack kill chain can vary depending on the adversary and target system. Furthermore, for zero-day exploits, no signatures exist and behavioral techniques such as machine learning must be employed. It is also necessary to consider cyber-physical indicators for systems such as the electric grid to understand when both cyber and physical corrective actions are necessary. Both cyber and physical data are not monitored concurrently either during planning or operational phases to detect when a cyber-physical event occurs. Therefore, adaptive and real-time analysis is needed to understand the cyber-physical violations that may occur during unpredictable disturbances and effectively deploy cyber-physical corrective actions.

It is important to note that the playbook-style traditional SPS/RAS with triggering condition and correction pairs that are determined through extensive simulation studies in the planning stage will always play a crucial role. HARMONIE-SPS aims to supplement these playbooks with a real-time, adaptive capability that can learn new triggering conditions and recommend mitigations that can be then added to playbook and/or used concurrently.

2.3. Automated Response

Many research efforts, from both industry and academia, have gone into improving the flexibility and dynamics of SPS implementation. In [10], an event-based method was proposed to enhance SPS that are created to address specific frequency and voltage instability issues. Using transient energy analysis, the conventional SPS implementation can be adjusted with flexible triggering thresholds [11], and also adaptive corrective actions [12]. To mitigate the risk of voltage instability and voltage collapse, BC Hydro developed a methodology to determine the magnitude of load shedding based on real-time measurement data [13].

Recent work [14] by the Pacific Northwest National Laboratory (PNNL) proposed an approach to adaptively set the arming parameters of existing SPS based on realistic and near real-time operation conditions. In collaboration with PacifiCorp and Idaho Power Company, a prototype named Transformative Remedial Action Scheme Tool (TRAST) was developed with advanced computing methods for adaptively setting SPS coefficients with the consideration of realistic and near real-time operation conditions. The Jim Bridger RAS, owned and operated by PacifiCorp, was used as the case study for testing and validating the methodology and prototype.

The TRAST tool performs statistical analysis for full-year supervisory control and data acquisition (SCADA) set provided by utilities that contain essential variables for the existing SPS model. Correlation analysis and regression is performed between these variables, as well as with temporal data such as season and month and power flow data from state estimator cases. A machine learning framework was developed to update the RAS coefficients; full details on the framework can be found in [14]. To summarize, this online SPS tool is automated in the sense that the parameters adapt to real-time conditions, however the design of the underlying SPS itself is manual. The other key factor is the vast amount of real system data (measurements and models) needed for the entire framework, spanning multiple entities and even years.

Overall, although the need for an adaptive SPS has been recognized, especially for the SPS development and triggering phases, adaptive and cyber-physical SPSs have not been proposed. The grid is increasingly cyber- physical and impact from either domain can easily propagate to the other - cyber-physical disturbances must be protected against for improved grid resilience.

3. HARMONIE-SPS DESIGN

In this section, an overview of the HARMONIE-SPS principles, subsequent design, and implementation are described.

3.1. Overview

For increased grid resilience, we hypothesize that an adaptive and reactive cyber-physical SPS is necessary for defending against diverse predictable and unpredictable disturbances, inadvertent or malicious. Therefore, we propose a defensive, wide-area SPS that learns system conditions, mitigates cyber-physical consequences, and preserves grid operation during both predictable and unpredictable disturbances.

HARMONIE-SPS will:

1. Detect and defend against cyber attacks that do not fit predefined abnormal conditions by using machine learning (ML) classification and anomaly detection algorithms,
2. incorporate intrusion detection system (IDS) and out-of-band (OOB) data for increased situational awareness, and
3. proactively respond to cyber-physical compromises by deploying distributed control algorithms and taking cyber-side actions (e.g., rejecting setting/firmware changes) to reduce and/or eliminate system impact.

An overview of the approach is shown in Figure 1. HARMONIE-SPS will prioritize selectivity, speed, and security using ML algorithms to classify system conditions as detailed in the next section.

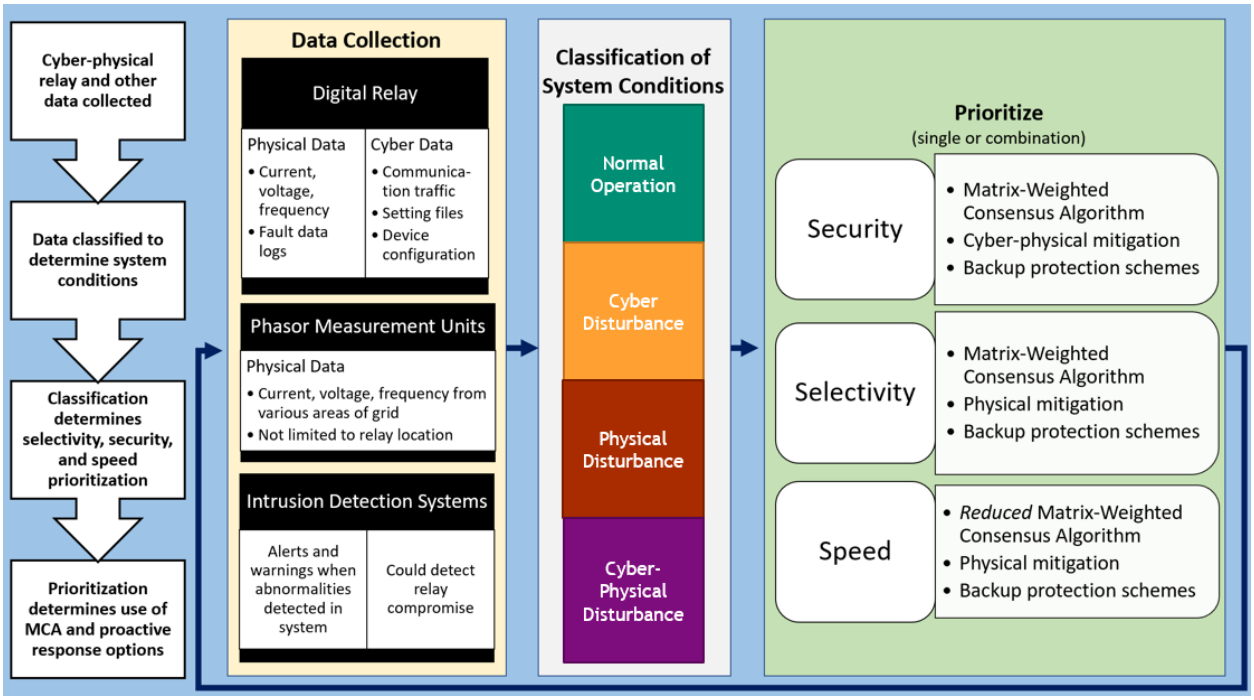


Figure 1: Overview of HARMONIE-SPS data collection, classification, and response prioritization.

3.1.1. Prioritizing Speed, Security, and/or Selectivity

Utilizing the classification of system conditions, HARMONIE-SPS will decide to prioritize selectivity, speed, or security; additionally, a combination could be selected. If a relay within a zone is compromised, selectivity will be enabled by deploying the consensus algorithm-based voting scheme approach that considers diverse zone relays and OOB data, as detailed in Section 5.3.

Unlike traditional relay voting schemes that simply compare tripping decisions, the proposed consensus algorithm-based voting scheme would account for inter-relay relationships in different zones and provide the ability to assign weights depending on the disturbance location and indication of specific relay compromise or failure (e.g., assign zero weight to that relay's vote). Thus, selectivity can be achieved by ensuring relay status and relationships are incorporated into the voting for high confidence in relay actions for the most up-to-date system conditions.

If detrimental system conditions are observed, speed will be prioritized by taking proactive actions such as switching to backup protection schemes to reduce impact propagation and/or deploying reduced-order voting scheme to achieve both speed and selectivity. Backup protection schemes could compensate for compromised/failed relay zones and provide increased protection to critical grid components such as generators and transformers.

When HARMONIE-SPS prioritizes speed, methods to quickly reduce/eliminate system impact and provide the most protection possible are needed. For both selectivity and speed, physical mitigations could also be deployed to maintain system operation. This could include distributed control approaches that employ power system devices such as distributed flexible AC transmission system (D-FACTS) and design their response to limit disturbance impact [15], [16]. Distributed decision-making algorithms, such as alternating direction method of multipliers (ADMM), could be explored as a powerful distributed convex optimization approach for making control decisions between distributed devices in response to system disturbances [17].

Lastly, security is enhanced with HARMONIE-SPS by 1) taking proactive actions to minimize impact propagation, 2) supplying relay data to augment analysis and aid IDSs in identifying the disturbance, and 3) providing confidence in tripping decisions with the novel consensus algorithm-based voting scheme. The proactive actions taken by HARMONIE-SPS include both cyber-side and physical-side mitigations. The physical-side mitigations encompass the distributed control approaches and switching to backup protection schemes whereas the cyber-side actions could include rejecting further firmware/setting changes, communicating relay compromise/malfunction to other peer relays, and restoring backup device configuration files. Additionally, if an IDS exists in the system, HARMONIE-SPS's findings on abnormal relay behavior (e.g., co-located, different zone relay measurements do not match) and classification results can be used to supplement the IDS data collection and analysis.

3.1.2. Evaluating HARMONIE-SPS

The high-fidelity cyber-physical emulation environment will be constructed using SCEPTRE™ and a real-time digital simulator such as RTDS. SCEPTRE™ is Sandia's industrial control system (ICS) modeling platform that enables modeling of different ICS devices (virtual and hardware) such as protective relays and programmable logic controllers, network components (e.g., gateways, switches, servers), actual ICS communication protocols (e.g., Modbus, DNP3, IEC 61850), and physical end processes (e.g., power system simulations) [18]. The RTDS will be used to model the physical end process and enable connecting protective relays both virtually and as hardware-in-the-loop (HIL) [19]. This emulation environment is discussed in Section 6.

In this manner, metrics can be collected to verify if HARMONIE-SPS reduced or eliminated system impact from a disturbance, malicious or inadvertent, and that network burden was not worsened with the application of HARMONIE-SPS (e.g., increased latency, dropped packets). Malicious cyber-physical disturbance scenarios are being developed to reflect a wide range of attacks and based on real-world events (e.g., using MITRE ATT&CK framework [20]).

4. HARMONIE-SPS CLASSIFICATION FRAMEWORK

The classification of system conditions for HARMONIE-SPS needed to 1) be able to process diverse types of datasets from both cyber and physical domains, 2) classify the system conditions into cyber, physical, cyber-physical, or normal events, and 3) have online, real-time capability. The next few subsections describe the machine learning framework that was developed to achieve these goals.

4.1. Integrative Modeling of Cyber-Power Data

One of the primary goals of the HARMONIE-SPS project was to develop a joint model of cyber-power information using an integrated data framework, enabling downstream prediction of anomalous events via machine-learning and other approaches. Although cyber-system and power-system data can both be modeled as a time series, developing an appropriate integrative data framework for these two data types is challenging, because the two types of data have very different temporal dynamics.

Cyber data is typically represented using a discrete event framework, in which individual packets or network flows are sent and received at irregular intervals in time. The discrete event framework is appropriate, as a given sender and receiver will typically only interact very sporadically over a specific port and protocol, leading to a very sparse interaction over time. Many properties of cyber network data are based on unique identifiers with no natural notion of 'distances' between identifiers, including the sender's and receiver's network IP addresses and port numbers, and the protocol over which the sender and receiver are communicating.

In contrast, power-system data is typically represented as a continuous time sequence, which naturally matches the information collected from a typical power system. Information about voltage, current and frequency is commonly sampled from various points in the power system using sensors that provide readings on millisecond timescales, resulting in a dense timeseries dataset of near-continuous variation. Power-system data can generally be represented as vectors of real-valued numbers having natural distance metrics.

Given that both cyber and power data can be represented as time series, we examined the "transformer" machine-learning framework using multi-head attention as a basic framework for integrating cyber-power information [21]. The transformer framework is a computation-efficient method for inferring information from data represented as a sequence, which is a basic data framework compatible with both discrete-event cyber data and continuous-time power data. Transformers are widely used in natural language processing and other sequence-based machine-learning problems, where they have largely superseded recurrent neural networks as state-of-the-art [22]–[24].
We implemented a transformer-based neural network for predicting cyber-power disturbance events from time-sequence cyber- and power-system data (Figure 2). Cyber-data and power-data were embedded into time-sequence vectors and integrated using either graph neural networks or modulated linear projections (see sections below). Integrated cyber-power time-sequence vectors were processed using a multi-layer transformer [21], [25], the output of which was evaluated by a linear classification head to determine the 'probability' of either a cyber- or power-disturbance event occurring at any point within the input time-sequence.

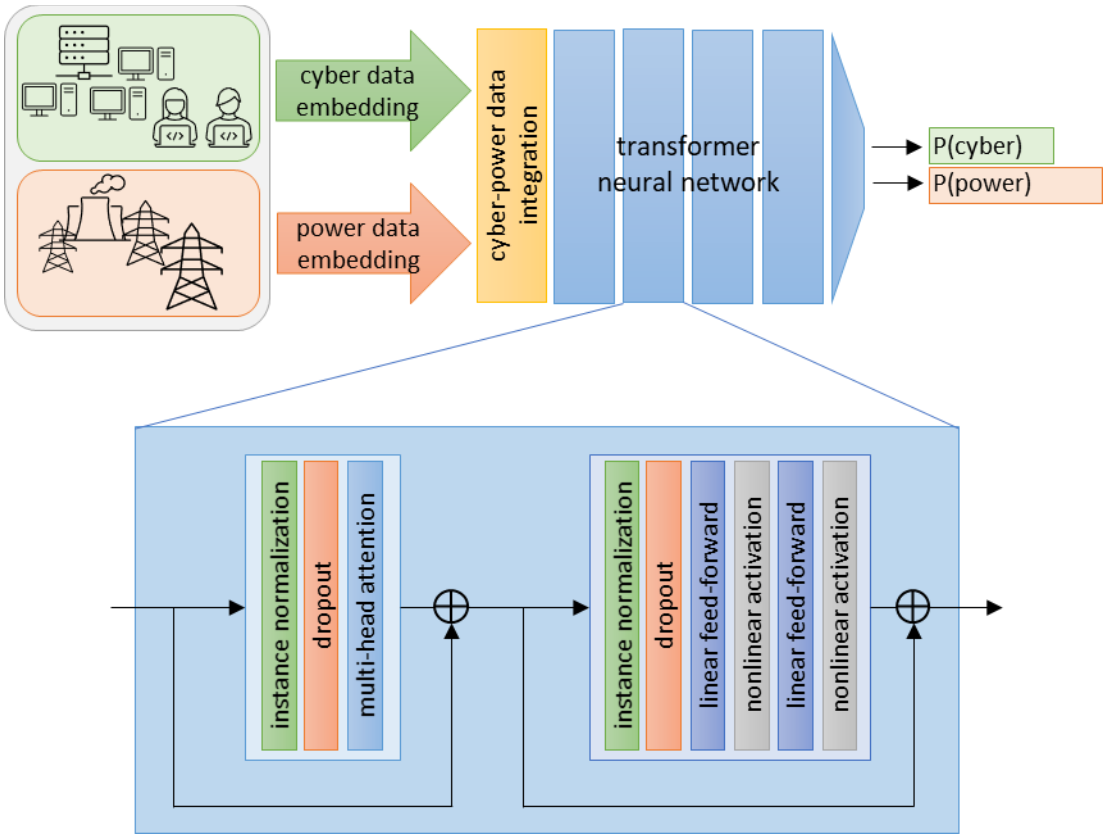


Figure 2: Transformer-based neural network architecture for predicting cyber- and power-disturbance events. Top: Data obtained from joint cyber-power system emulations is embedded into a time series sequence using various methods (see below) and integrated for neural-network analysis using a multi-layer transformer network, which is trained to predict the 'probability' of either a cyber-disturbance or a power-disturbance event occurring within the time sequence. Bottom: Each transformer block in the network consists of a multi-head attention sub-block (left), followed by a sub-block of 2 feed-forward linear layers with nonlinear activations (right). We used pre-instance-normalization and dropout (p=0.2) before each sub-block and used residual connections around each sub-block (indicated by the circled cross) [18], [22].

A major limitation of the transformer framework is that multi-head attention—although computation-efficient—is memory-*inefficient*; the naïve implementation requires $O(n2)$ memory, where n is the sequence length. Practically, this limits transformers to processing sequences of length <512 on most common graphics processing unit (GPU) hardware. Approaches to address the memory-inefficiency of transformer-based neural networks have been proposed, including 1) combining sliding-windows with global attention [26], 2) grouping similar datapoints in the sequence together and computing multi-head attention only within data-similar groups [27], 3) approximating the large matrix multiplications required for multi-head attention using memory-efficient methods [28] and 4) randomizing attention mechanisms across long sequences [29].

In one study, we found that when time-series sequences were long, a sliding-window approach using time-local multi-head attention was slightly more accurate at detecting cyber disturbances (0.98 Area Under the Curve (AUC) vs 0.95 AUC for randomized attention), whereas randomized attention was slightly more accurate at detecting power disturbances (0.87 AUC vs 0.85 AUC for time-local attention). Although absolute differences in performance were small (<0.03 AUC), these results may

suggest that longer-scale dependencies in time-sequence data may be more important for accurately predicting power-system disturbances, compared to disturbances in the cyber-system [30].

In addition to reducing the memory requirements of the transformer framework by modifying the multi-head attention mechanism, the transformer's memory requirement can also be reduced by compressing the sequence data to produce a shorter sequence. To investigate this approach, we used a partially-overlapping time-window approach that averaged cyber-power data samples over a 1-second time window to produce a time sequence over ~30 seconds (the same time sequence used in our previous study [30]). Specifically, we slid a 1-second averaging window by 0.5-second intervals over 32 seconds of cyber-power data capture, producing a sequence of 64 time samples, each averaged over 1 second. Data were generated using RTDS emulation of the WSCC 9-bus system under normal system operation and 3 different disturbance events: 1) a denial-of-service cyber disturbance, 2) a line-outage power disturbance, and 3) combined denial-of-service and line-outage disturbances (cyber-power disturbance).

We trained a 283,292-parameter neural network to predict the presence of a cyber, power or cyber-power disturbance within each 32-second data capture from the sequence of 64 partially-overlapping 'averaged' time windows. The network consisted of 8 transformer layers, each using a typical residual multi-head attention layer with 4 attention heads, followed by a residual sequence of 2 position-wise feed-forward layers [21] (see Figure 2). We used latent-space encoding vectors of 64 dimensions for both cyber and power data, with instance normalization and dropout ($p = 0.2$) before each residual path. We did not use bias neurons in multi-head attention or feed-forward layers, and we applied a novel activation function after each feed-forward layer, which was calculated as:

$$y = (tanh(x) + (0.6 * x)) * 0.6$$

This smooth activation (which we called "unbounded-tanh") is centered at zero, roughly linear with slope ~ 1 when $-0.6 < x < 0.6$, and roughly linear with slope 0.4 when $x < -2.5$ or $x > 2.5$. Intuitively, unbounded-tanh mimics the traditional tanh activation function for small input values x but produces approximately-constant non-zero gradients as x becomes very large (either positive or negative), avoiding the vanishing-gradient problem associated with tanh activation, the gradient of which approaches zero as grows < -2.5 or > 2.5 [31].

The network's decision layer consisted of a linear projection (without bias) from the flattened transformer output to a 2-dimensional vector (after dropout of 0.2), with the first dimension being the logit-score for a cyber disturbance event, and the second dimension being the logit-score for a power disturbance event. Sigmoid activations were used respectively on the logit scores to calculate P(cyber) and P(power) disturbance 'probabilities'.

We trained the transformer model for 200 epochs using 600 emulated data samples (batch size 64; 58 samples with cyber-disturbance, 60 with power-disturbance, 38 with both cyber- and power-disturbance, and 444 with no disturbance) and binary cross-entropy loss, using the AdamW optimizer with $beta\_1 = 0.9$ and $beta\_2 = 0.99$ and $weight\_decay = 0.01 * sqrt(batch\_size / data\_samples * epochs)$ [32]. We increased the weight of each positive training sample (cyber or power disturbance) by the number of negative samples (normal operation) in the training data set divided by the number of total positive samples. We used a one-cycle learning rate schedule with a maximum learning rate of 0.001, an initial learning rate of 0.04 times the maximum learning rate, and

a final learning rate of 0.0001 times the maximum learning rate [33], [34]. We used cosine learning-rate annealing, and we varied momentum between 0.85 and 0.95 during training (see Figure 3 for experiment overview). Model training was replicated 4 times using random stratified splits of the 600 data samples into 80% training and 20% validation data. Results are presented using validation datasets.

To address the relatively small amount of training data, we used both data augmentation [35] and adversarial training approaches [36]. Although data augmentation has been widely studied for image data, sequence data augmentation has received much less research attention and, to our knowledge, data augmentation has not been applied to the cyber- and power-system data examined in this study. We developed novel data augmentations based on the RandAugment framework developed for image data [35]. Briefly, we developed novel augmentations that randomly permute either the sender and receiver (for cyber data) or the PMU source (for power data), as well as a widely-used data augmentation that smooths the data labels [37]. For each data sample, we randomly selected 2 data augmentations with replacement (including the 'null' augmentation that does not alter the data or labels) and applied those augmentations before training on the data sample. In the case of the label-smoothing augmentation, we additionally randomly selected the 'strength' of the label smoothing (between zero and 0.1) before each augmentation.

For adversarial training, we used a combination of projected gradient descent (PGD, [38]), a randomized variant of the fast gradient sign method (FGSM, [39], [40]), a novel approach that modifies k/n elements in the time sequence, and a 'mixin' adversary that generates linear interpolations between two randomly-selected data samples. In addition to taking the sign of the gradients, we implemented novel variants of PGD and FGSM that use normalized gradients. For each training batch, we selected an adversarial-training algorithm at random and randomly-selected the 'strength' of the adversarial examples (typically the epsilon-budget) from between zero and 0.2, scaled to the maximum value of the data samples. In the case of PGD-based adversaries, we used 16 steps to generate each adversarial example. We applied the chosen adversarial algorithm to every sample in the training batch and appended the adversarial examples to the original training batch during training. Note that adversarial training was applied after data augmentations on each training batch.

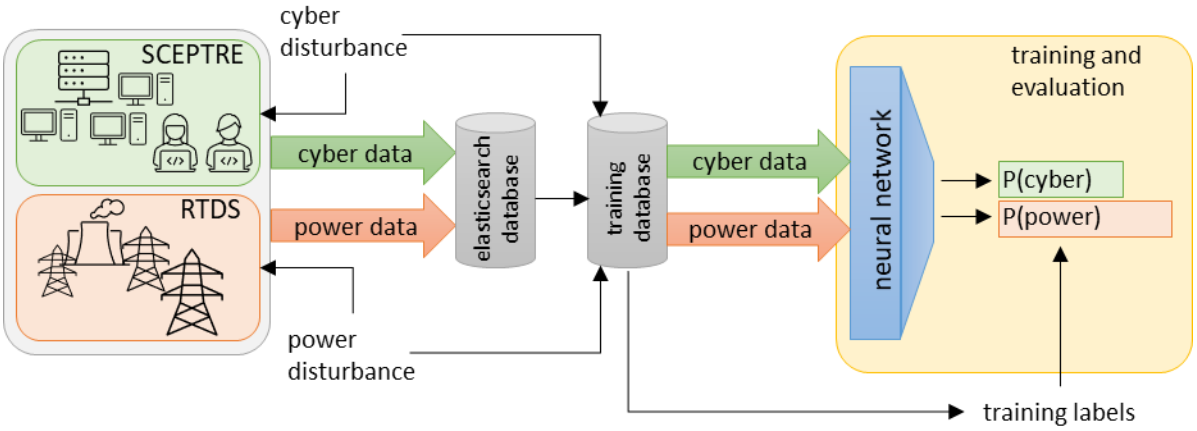


Figure 3: Experimental protocol for training cyber-physical disturbance classification network. We simulated joint cyber-power systems using SCEPTRE and RTDS under a variety of disturbance scenarios (see main text for details). The resulting cyber (green) and power (red) data streams were stored in an Elasticsearch database, which was used (along with 'ground truth' disturbance labels) to construct a neural-network training database consisting of 32-second data captures over the course of each simulation. Training data were labeled as a cyber-disturbance if a cyber-disturbance event occurred at any time within

the 32-second window (similarly for power-disturbance events). The time-windowed training data and labels were used to train a neural network to classify time-windows independently as cyber-disturbances or power-disturbances (or both; see main text for training details).

We found that, in the case of the WSCC 9-bus system with either denial-of-service cyber-disturbance, line-outage power-disturbance, combined cyber-power disturbance or no-disturbance, the transformer-based neural network model achieved near-perfect validation accuracy when trained with strong data-augmentation and adversarial examples (Figure 4). While these results are encouraging, we caution against over-interpreting these positive findings due to the small size of the training dataset (600 examples), lack of experimental replication (each simulation experiment was only run once, and the network was only trained once), and lack of independent testing dataset. Further experimental validation is clearly necessary to investigate the reliability and generalizability of these results.

With these cautions in mind, we found that training our transformer network using data from the RTDS-SCEPTRE emulation system produced near-zero false-positive (incorrectly classifying a no-disturbance data sample as having either cyber- or power-disturbance) and false-negative (incorrectly classifying a cyber-disturbance, power-disturbance or cyber-power-disturbance sample as having no-disturbance) error rates on the validation dataset. By the end of training, we did observe elevated false-positive and false-negative error rates on the training dataset sample with data augmentation and adversarial examples (e.g., 8.9% of training samples without power-disturbance were classified as having power-disturbance (power-fp); the cyber-disturbance false-positive rate was 5.8% on training samples; power-disturbance false-negative rate was 24.8%, and cyber-disturbance fn rate was 25.5%). But error rates on the validation dataset were always $< 1e-10$ (note that validation data were sampled without data augmentation or adversarial examples). In this case, we used a 50% probability cutoff to calculate false-positive and false-negative error rates (see Figure 4).

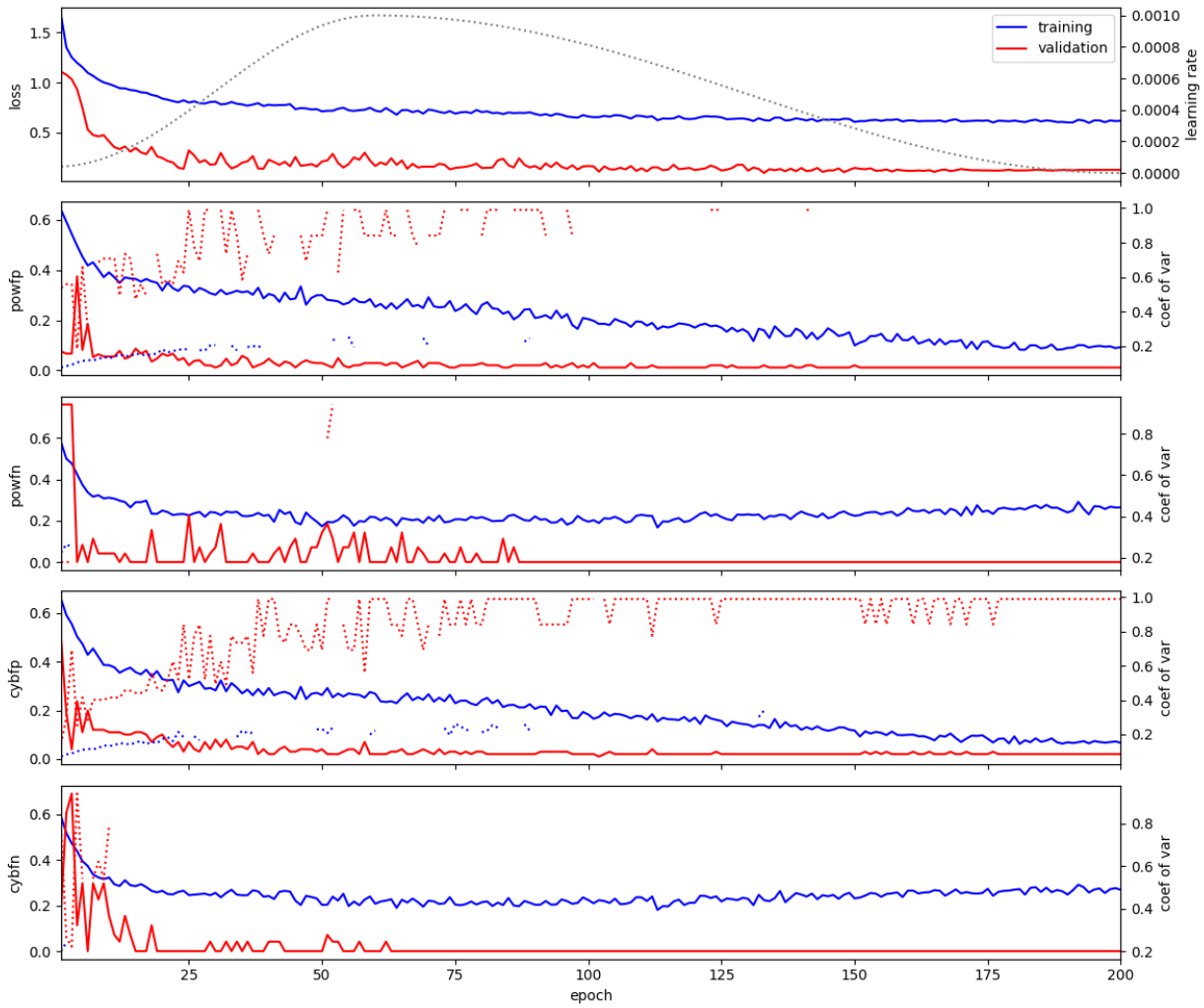


Figure 4: Transformer neural network produces low false-positive and false-negative rates on validation dataset when trained with data augmentation and adversarial examples. We trained a transformer-based neural network using emulated data collected over 32-second intervals with cyber-disturbance (cyb), power-disturbance (pow) or both cyber- and power-disturbance (see main text for details). We plot the loss, power-disturbance false-positive and false-negative error rates (powfp, powfn, respectively), cyber-disturbance false-positive and false-negative error rates (cybfp, cybfn, respectively) and learning rate (dotted line, top) averaged over all data batches in each epoch of training. Blue lines indicate training data, and red lines indicate validation data. Coefficients of variation (coef of var) are plotted as dotted series but are not visible when they are near-zero. Results are shown for one replicate training run; results from other replicates are in Appendix A.

It is interesting to note that the false-positive and false-negative error rates were both roughly balanced across cyber-disturbance and power-disturbance events. Data samples with either cyber-disturbance or power-disturbance events were misclassified as no-disturbance (a false-negative) roughly 25% of the time when data-augmentation and adversarial-examples were used during training. Although the false-positive rate for cyber-disturbance events (misclassifying a no-disturbance or power-disturbance as a cyber-disturbance) was slightly lower than that of power-disturbance events (5.8% vs 8.9%), the difference was relatively small in magnitude, and false-positive error rates were always lower than false-negative error rates on the training data. We note that modifying the positive-training-sample weights during training and/or modifying the probability cutoffs used to classify false-positive and false-

negative errors is expected to strongly impact these calculations and could be used to fine-tune network performance in practice, particularly if independent testing data were available; we leave these exercises for future work.

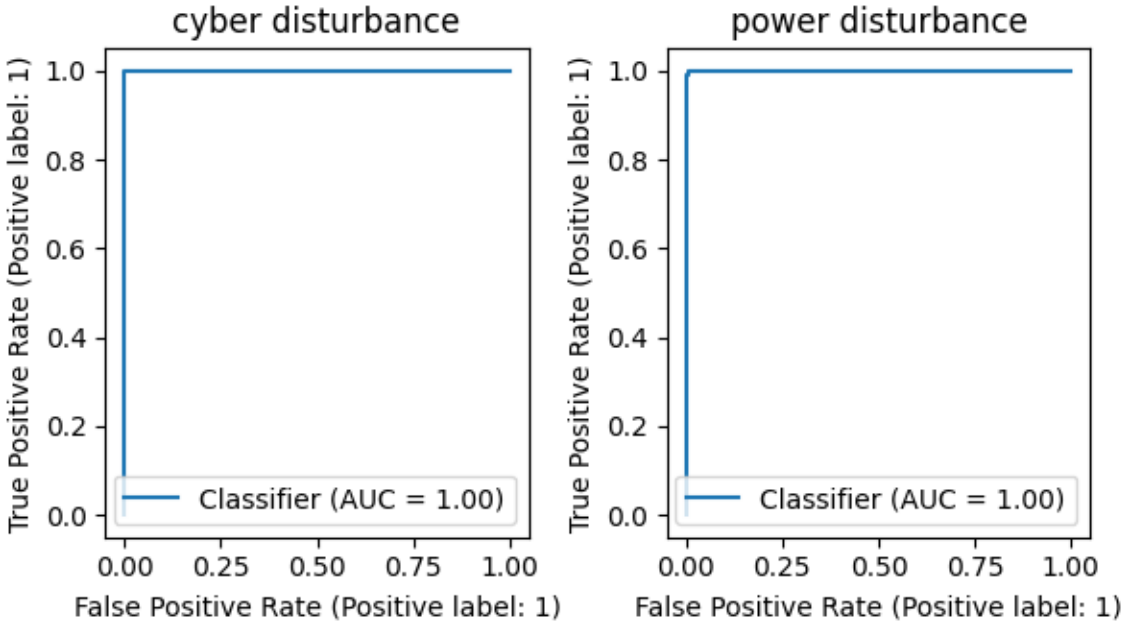


Figure 5: After training, transformer model validates with perfect accuracy on cyber- and power-disturbances. We plot ROC curves for identification of cyber (left) and power (right) disturbances using the trained transformer model (see Figure ML03). Results are shown for one replicate training run; results from other replicates are in Appendix A.

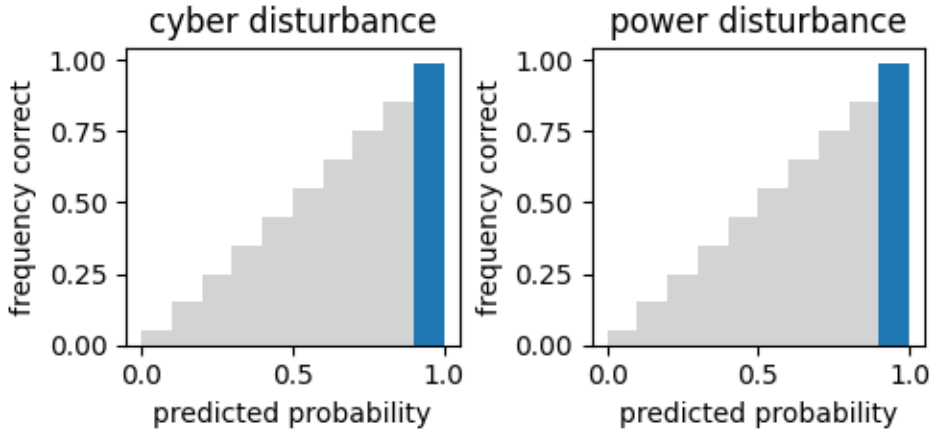


Figure 6: Trained transformer model infers cyber- and power-disturbance events with high statistical confidence. We plot the predicted probability (x-axis) of cyber (left) or power (right) disturbance event vs the frequency with which events of the indicated predicted probability were correct (y-axis). Predicted probabilities were binned every 0.1, and bins with <10 samples were removed. Gray bars indicate 'ideal' frequentist probability distribution. Results are shown for one replicate training run; results from other replicates are in Appendix A.

We verified the exceptionally high accuracy of the transformer model under these cyber- and power-disturbance scenarios after training by calculating receiver operator curve (ROC) curves (Figure 5). As expected based on the training results, the transformer model performed 'perfectly' on the validation data, with all true-positive examples being correctly identified before any false-positives (AUC = 1.0). These results were consistent for both cyber- and power-disturbance events (including time windows with both cyber- and power-disturbances).

Finally, we used the sigmoid function ($y = 1/(1 + exp(-x))$) to transform the model's output logit scores into an approximation of the 'probability' of either a cyber- or power-disturbance event. Under standard 'frequentist' expectation, the inferred probability of a disturbance event should match the large-sample frequency with which the event occurs over many trials. To test whether the model's inferred probabilities match this expectation, we binned probability scores every 0.1, calculated the proportion of validation data samples within each bin that were correct, and plotted the results as a histogram (Figure 6). If the model's inferred probabilities precisely match frequentist expectations, approximately 0.95 of the data samples with $0.9 < probability < 1.0$ should be correct, ~0.85 of samples with $0.8 < probability < 0.9$ should be correct, etc. In our experiment, we found that all model inferences had probability equal to 1.0, and all the model's inferences were correct. While this result trivially matches frequentist expectation, unfortunately there was not enough variation in the model's probability output to reliably evaluate. Future experiments using scenarios with more challenging data will be required to investigate the frequentist properties of the model's probability outputs.

4.2. Integrating Cyber-Power Data with Graph Neural Networks

Both cyber-communication and power-distribution systems can be represented in a computer as a network or 'graph' consisting of 'nodes' connected to one another by 'edges'. In the case of a cyber network, nodes would naturally represent computers connected to the network, with edges representing direct physical connections among the networked computers. The power-distribution system could similarly be derived by mapping power generation/consumption devices to 'nodes' and distribution lines/busses to 'edges'. However, this approach has two primary limitations: 1) it fails to capture the mechanism by which power-system data are obtained in practice, and 2) it does not provide an integrative mapping between cyber- and power-networks.

To address these issues, we developed an integrated cyber-power network representation using power-monitoring hardware devices (phasor measurement units or PMUs) as a natural 'connection' between the cyber-communication and power-distribution systems. In our integrated cyber-power network model, we build a graph model of the cyber-communication network, with PMU nodes added to the graph and injecting data as edges between each PMU node and the power-system's control center node (Figure 7).

The integrated cyber-power graph model is used as a topology to support 'message-passing' data integration via a Graph Convolution neural Network (GNN, [41]). Briefly, a GNN uses an iterative approach to combine information across nodes(edges) in a graph, given the graph topology. The GNN begins with a learned information representation (randomly initialized before training) at each node(edge) in the graph topology. At each step in the iteration, data at adjacent nodes(edges) in the graph is integrated into the node's(edge's) information via a neural network (of arbitrary type and complexity). We used 4 GNN iterations in our experiment. After the GNN iterations, each node(edge) in the graph contains an updated information representation, incorporating information from

nodes(edges) 'iterations'-steps away, given the network topology (iterations=4 in our case). Finally, we used a weighted linear projection to combine information across all nodes(edges) in the graph into an embedding vector time-sequence suitable for further processing using a transformer neural-network (see Figure 7).

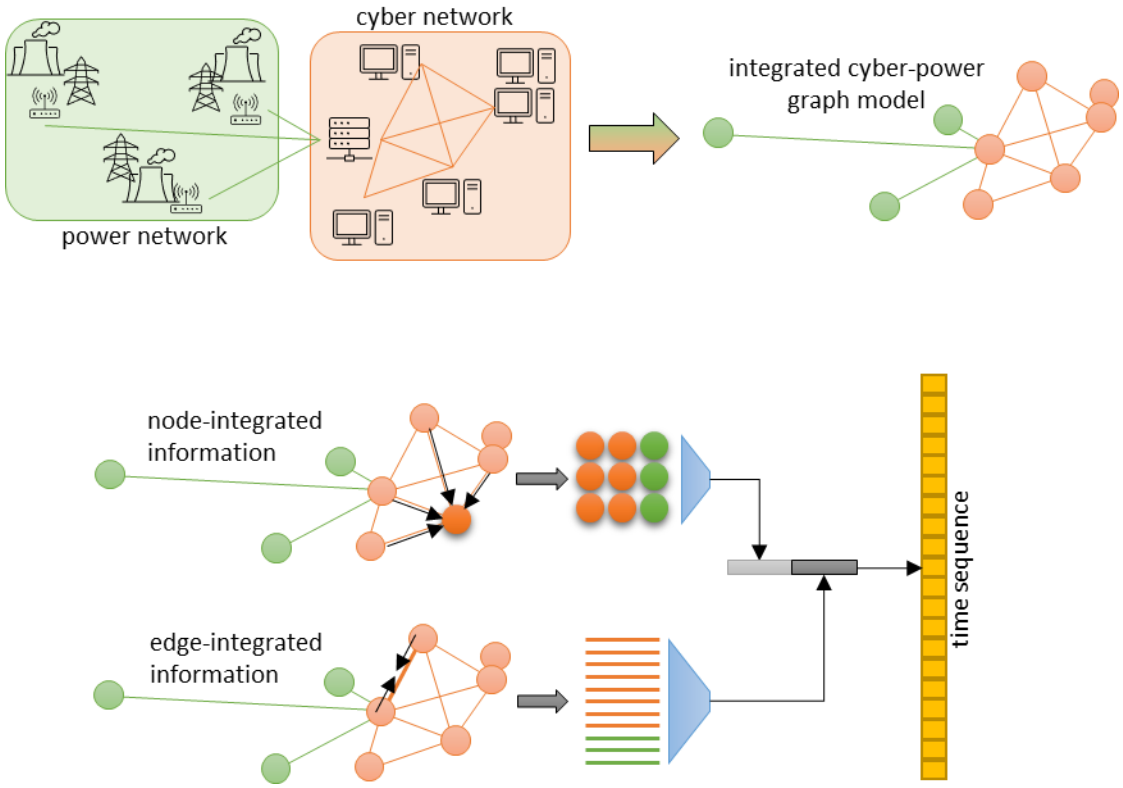


Figure 7: Graph Neural Network (GNN) approach for integrating cyber and power network data. Top: PMUs in the field transmit information about the state of the power network to the control center, which is networked with and communicates with other computers in the cyber network. We combine the power network (green) and the cyber network (red) topologies into a single integrated graph model representation (left). Bottom: The integrated cyber-power graph model is processed using a graph-convolution neural network (GCN), which iteratively combines information from adjacent nodes (top) or edges (bottom) in the graph [38]. Note that, for clarity, we highlight either a single node (top) or edge (bottom); in practice, the GNN updates every node/edge in the graph simultaneously. After a specified number of GNN iterations, the node (top) and edge (bottom) information is linearly projected to embedding vectors (gray/black), which make up a time sequence (orange) that can be processed by a transformer neural network (see Figure 2).

One potential advantage of our graph neural network approach is that it provides arbitrary-precision time-sequence information; that is, the approach uses a discrete-event model, in which each item in the time-sequence has a floating-point timestamp indicating precisely when the item occurred. This means that the time-sequence is 'dense' and naturally integrates both cyber data (which is naturally discrete-event) and power data (which is inherently discrete-time). A potential downside to this approach is that the time-sequence data structure can become too large to process using a standard multi-head attention transformer on common GPU hardware, necessitating the implementation of a multi-head attention 'approximation' with lower memory consumption.

In our experiments [30], we found that different approaches to multi-head attention (MHA) approximation – although performing similarly overall – may require some 'trade-off' between accurate prediction of cyber-disturbance vs power-disturbance events. Specifically, a sliding-window approximation of MHA provided slightly more accurate identification of cyber-disturbance events, whereas a random-windowed MHA approximation was slightly more accurate at identifying power-disturbance events (although differences in AUC were <3%).

4.3. Integrating Cyber-Power Data with Modulated Linear Projections

As a potentially complementary approach to the graph-neural-network integration, we also investigated a modulated linear projection method for embedding cyber and power data into a shared latent vector space. Unlike the GNN approach, modulated linear projections completely ignore the network or graph topology of the cyber-power system and implement entirely independent latent-space projections for cyber and power data streams (Figure 8). For the power-system data, phasor data streams from each PMU were averaged over a 1-second sliding-window and stored in an NxM matrix, where N is the number of PMUs, and M is the dimensionality of the PMU data stream, which was 14 in our case (3 phases, voltage and current for each phase, magnitude and angle for each voltage or current measurement (=3*2*2=12 dimensions), plus the PMUs frequency and frequency-gradient information (=12+2=14 dimensions)). To embed the NxM matrix into a vector latent space, we first performed a learned linear projection of the 14-dimensional PMU data stream into a K-dimensional latent space (K=64 in our case). Next, we 'modulated' the initial linear latent-space projection via a second linear projection that applied a unique weight (and bias) vector(s) to each of the N PMUs, resulting in a final K-dimensional latent vector representation of each time window.

Embedding the cyber-network data into a vector latent space is slightly more complicated but follows the same general procedure as for power-system data (see Figure 6). In the case of cyber data, we ignore the network topology and instead capture the number of communication packets sent between each 'source' and each 'destination' node in the network. We watch a number of commonly-used network communication protocols and also include 'unknown' entries for both computer IP addresses and network protocols, allowing us to capture unusual network communication data. For each 1-second window of packet capture, we store the data in an $NxNxM$ 3-dimensional tensor, where N is the number of network nodes (+1 for unknown), and M is the number of watched communication protocols (+1 for unknown). Similar to what was done for the power-system data, we first project the M-dimensional packet count data into a K-dimensional latent space using a simple linear neural-network layer with K output units; this generates an $NxNxK$ latent-space tensor from the $NxNxM$ input. We next 'remove' the 'destination' dimension using a linear neural-network layer with a single output unit applied to the transposed $NxKxN$ matrix, which produces an NxK embedding matrix. Finally, we remove the 'source' dimension using the same approach: $NxK -> KxN -> K$, producing the final K-dimensional embedding vector for the cyber data. In our experiments, we simply concatenated the independent power-data and cyber-data embedding vectors to produce the combined cyber-power data embeddings.

This modulated linear projection was efficiently implemented using linear neural-network layers and tensor transposition. That is, we first processed the NxM PMU-data matrix using a linear neural-network layer with K output units to produce an NxK matrix: $NxM -> linear -> NxK$. We then transposed the NxK matrix into a KxN matrix and processed this data using a linear neural-network

layer with a single output unit to produce an embedding vector of dimension K for each 1-second time window: $NxK -> KxN -> K$.

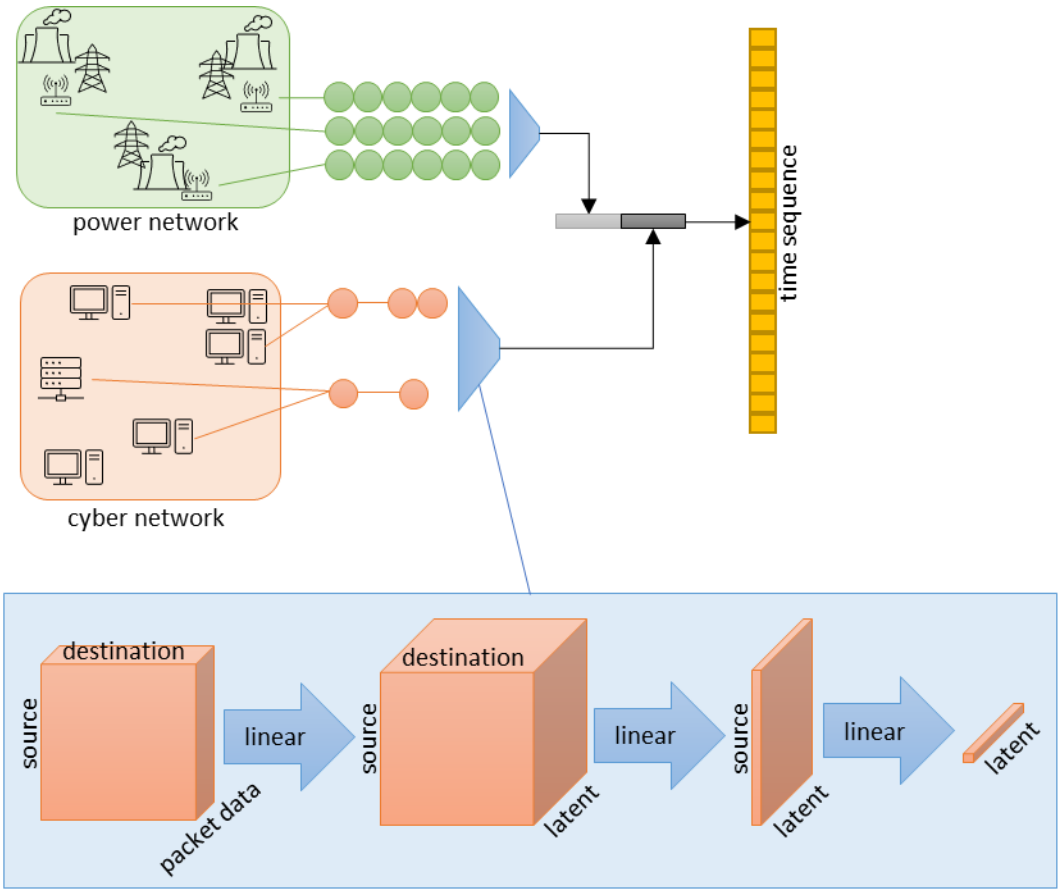


Figure 8: Modulated linear projection approach for embedding cyber and power data. Top: We ignore network topology and capture each PMU's data stream (top, green) and communication packets sent between each pair of computers on the network (bottom, red). Information is averaged over a 1-second sliding window. Power and cyber data streams are processed independently using modulated linear projections (blue) into latent space vectors at each 1-second window in the time sequence (orange). Bottom: We show the modulated linear projection for cyber data (power data projection is described in main text and is similar). For each 1-second time window, we count the number of packets sent from each 'source' node in the network to each possible 'destination' along each watched network protocol. We project the packet-data tensor dimension into a K-dimensional latent space using a linear projection (implemented as a linear neural-network layer). Next, we use a second independent linear projection to 'collapse' the 'destination' dimension to 1, creating a 2-dimensional matrix of SOURCExLATENT. Finally, we use a third linear projection to remove the 'source' dimension and embed the 3-dimensional input tensor into a latent-space vector.

Our modulated linear projection method was derived from StyleGAN, which uses a similar method of applying a sequence of learned linear projections to either data or network weights in order to project 'style' information and control image-generation processes [42], [43], and from depthwise-separable convolutions, which use a similar dimensionwise-decomposition approach to reduce model complexity [44], [45]. To our knowledge, ours is the first application of this technique to latent-space embedding of high-rank input data. It is more parameter-efficient than projecting high-rank input

tensors directly into vector space using a fully-connected linear layer, while providing some capacity to adapt the embedding process differently to each input rank.

One potential advantage of the modulated linear embedding approach is that it allows the length of the time sequence to be controlled by changing the time-window size and the amount of time over which data is captured. By averaging data over a sliding-time-window, the resulting time-sequence can be kept short enough to fit in GPU memory, allowing standard global multi-head attention libraries to be used to construct the transformer neural network. A potential downside is that averaging over a sliding window inherently loses information, and the sparse nature of cyber-network communication packets over time could impact neural-network performance under some scenarios.

Although our preliminary results suggest that this approach may provide highly accurate disturbance predictions in some cases (see above), further experiments are required to validate this proposed approach. For example, it is unclear how this approach may perform across a much wider variety of disturbance scenarios and/or using larger and more realistic cyber-power network simulations. Determining the best approaches for integrating the results of ML predictions into operator or automated SPSs is also an important area for further research.

5. HARMONIE-SPS CYBER-PHYSICAL MITIGATIONS

For HARMONIE-SPS, a primary focus is to demonstrate the need of having both a suite of cyber mitigations and physical mitigations to be able to effectively respond to unpredictable (and predictable) disturbances adaptively. The cyber and physical corrective actions we considered were not exhaustive of all the different kinds of actions that can be deployed but focused on how both can be necessary for certain disturbances. A few novel mitigations were developed, including a routing optimization algorithm for network traffic, the automated corrective action and triggering condition pairing through autoRAS, and the consensus algorithm-based relating voting. More details on these novel mitigations are provided in the remainder of this report. A summary of the corrective actions explored during the duration of this project are listed below.

- **Cyber Mitigations**
    - Updating firewall rules
        - Blocking attacker IPs
        - Dropping malicious traffic
    - Developed traffic rerouting optimization algorithm
- **Physical Mitigations**
    - Line switching, load shedding, changes to generator setpoint, etc.
    - Developed autoRAS approach that uses Support Vector Machines to automatically assign corrective actions to triggering conditions based on clustered violations
- **Cyber-Physical Mitigations**
    - Combination of above listed individual mitigations
    - Relay voting that incorporates cybersecurity information (e.g., compromise alert)

5.1. Cyber-Physical Mitigation Testing within Emulation Environment

To test the detection of cyber-physical disturbances and mitigation impact, a few different experiments were performed in the emulation environment with the WSCC 9-bus system described in Section 6. These experiments are described in Table 1.

Table 1: Cyber-Physical Disturbance and Mitigation Experiment Description

Disturbance Class	Disturbance Type	Disturbance Scenario	Mitigation
Cyber	DoS via DNP3	DoS from external attacker; compromise host/node detected	Block all attacker communications using firewall rules
Physical	Loss of Generator and Branch	Loss of generator and branch leads to overloading on two branches	Perform load shedding at two buses (amounts computed by autoRAS)
Cyber-Physical	DoS interrupts grid controls	DoS interrupts load shedding commands after generator and branch losses	Block all attacker communications using firewall rules and allow load shedding command to transmit

Each of the disturbances were implemented within the environment with the WSCC 9-bus system and each were correctly classified using the HARMONIE-SPS machine learning framework. More details on the classification results are provided in Section 4.

For the cyber disturbance, the mitigation to block attacker communications eliminated the impact on the round trip time (RTT) and allowed the communications to resume. This can be seen in Figures 9-12. With no DoS, it can be seen that the RTTs are similar with and without the firewall rule in Figure 9. Figures 10 and 11 show that the DoS causes a high amount of 6-7ms RTTs; Figure 12 shows that with the mitigation of the DoS, the RTTs are very similar to the nominal distribution and the mitigation is able to restore the system successfully.

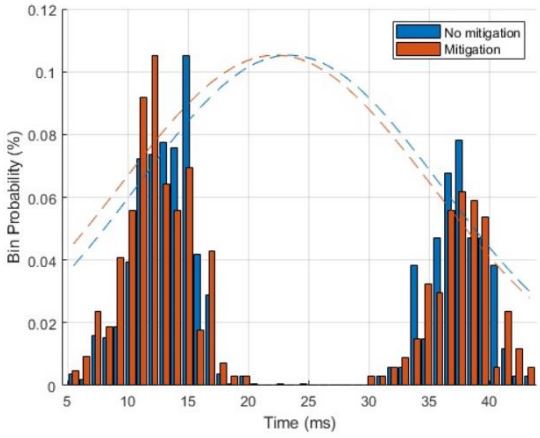


Figure 9: Nominal RTTs with no DoS.

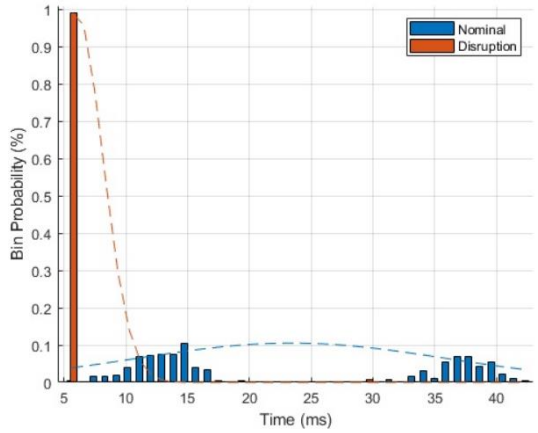


Figure 10: Nominal RTTs compared to DoS RTTs.

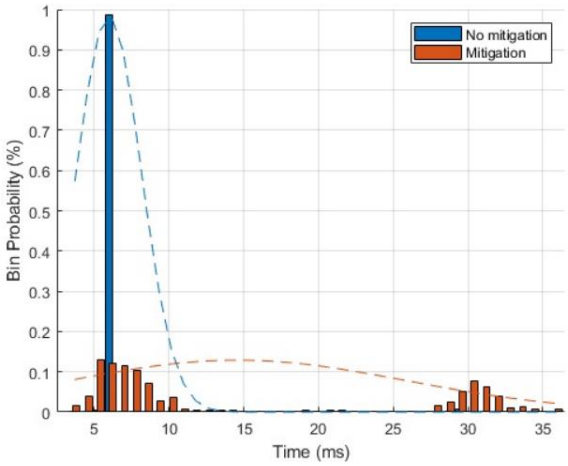


Figure 11: RTTs during DoS with and without mitigation.

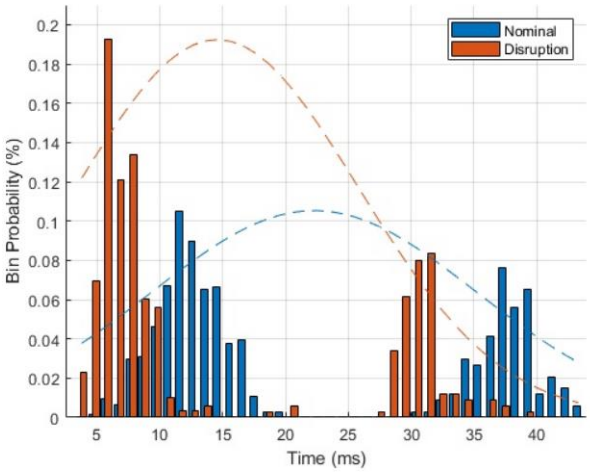


Figure 12: RTTs with mitigation.

For the physical disturbance, the loss of generator and branch caused overloading, which impacted the system frequency. The load shedding scheme was determined using autoRAS, described in the

next section, including the identification of the bus and the amount of load shed. The load shedding eliminated the branch overloads and restored the power system, demonstrating the effectiveness of the physical mitigation. Finally, for the cyber-physical mitigation, the DoS was mitigated using the firewall rules which then enabled the transmission of the load shedding command for mitigating the power system consequences. All in all, these experiments demonstrated the effectiveness of having a suite of both cyber and physical corrective actions to fully restore the system after disturbances.

5.2. autoRAS Approach: Automating Assignment of Triggering Conditions and Corrective Action Pairs

Traditional SPS/RAS design currently takes a holistic approach that requires years of experience with the operations of a specific system [9]. New RAS are proposed to reliability coordinators individually by entities such as transmission owners, generator owners, and distribution providers based on their operational experience. Each RAS is modeled and introduced to the system individually to address some predetermined contingencies that are known to cause violations of reliability or stability standards. To mitigate the potential violations or instabilities, numerous offline simulations are repeated on those predefined scenarios to ensure that the candidate remedial action is sufficient and does not introduce unintentional risks to the system. The implementation and testing of the RAS then involves manually defining the corrective action for a contingency definition within the energy management system (EMS) or other simulation software. The parameters associated with each scheme, such as the arming conditions, triggering threshold, and the numerical values of generation tripping and load shedding, usually do not change during real-time operations [8], [9].

In addition to the slow design process, the constantly evolving nature of the grid is compelling RAS to keep up with the changes. With the increasing penetration of distributed energy resources such as solar PV systems and wind farms, grid-edge devices, and the rise of unpredictable disturbances, a slow and manual RAS identification design process may not suffice. Also, to mitigate the impacts of these unpredictable events as they occur during grid operations, it is important to address the design with online timeframes in mind The corrective actions need to be computed fast, within seconds or minutes, depending on the type of violation. The power industry is cognizant of these evolving needs and hence there are regular assessments of the existing RAS' adequacy and the need for upgrades. For instance, in 2018 Hydro One in Canada replaced the RAS at their Bruce substation [13]. This increased the functionality to detect and operate for more contingencies and configurations, and enhanced system operation. Similarly, the need for and installation of new RAS are announced regularly by different entities, as presented in [46]. However, the design process still involves repeated, offline simulations.

Hence, the autoRAS approach was develop that provides a more flexible, computationally efficient approach to determine RAS corrective actions. The autoRAS design procedure aims to automatically generate triggering condition and corrective action pairs based on the identified need for new RAS creation. Statistical and functional characteristics summarized from RAS implemented in real power systems are utilized as a reference to guide the design parameters; sensitivity factors are also used to quickly design corrective actions for severe contingencies [47]. This automated RAS designed procedure is summarized in Figure 13.

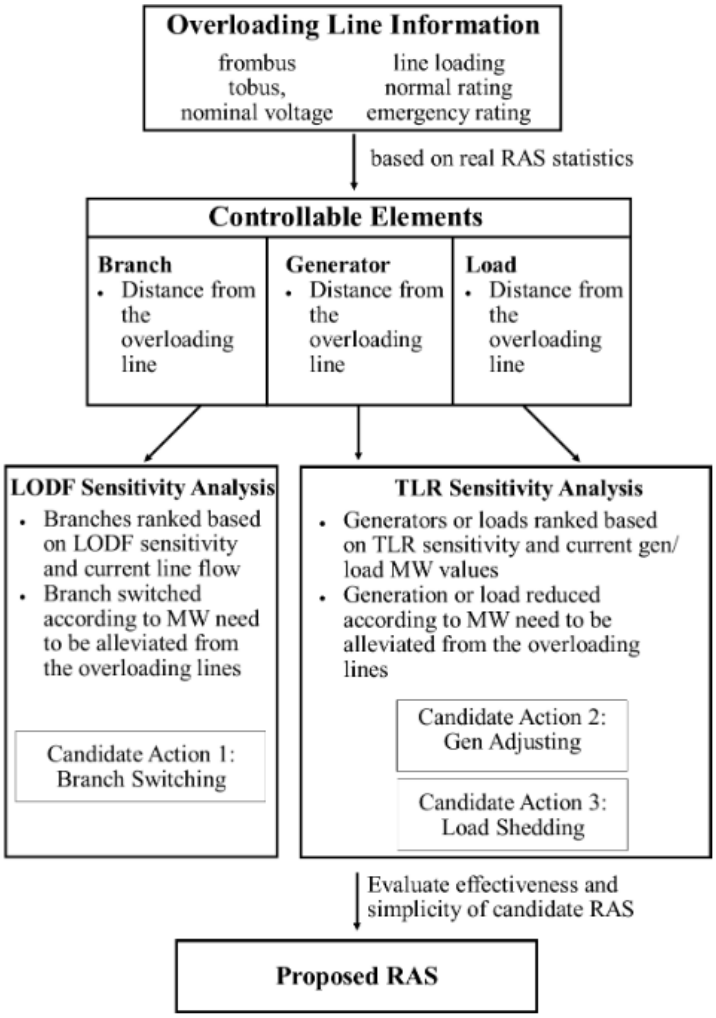


Figure 13: autoRAS approach leveraging system sensitivities to automate the assignment of triggering condition and corrective action pairs [46].

Further details and results demonstrating the use of the autoRAS approach with various use-cases are provided in [46], [48].

To extend the pairing of triggering conditions and corrective action pairs, the goal is to link the HARMONIE-SPS event classification to identify when a triggering condition has been encountered. An approach was developed to cluster (using hierarchical clustering algorithm) violation elements resulting from contingency analysis applied to a use-case -- the WSCC 9-bus system (22, 000 scenarios were generated and a subset was sampled). The violations within the same cluster can then be addressed with one corrective action, as identified using a support vector machine (SVM) approach. An example of this implementation is shown in Figure 14 and 15. Ultimately, the outputs of the SVM will be inputted into the GNN and RNN framework to confirm the corrective action in future iterations of HARMONIE-SPS. Foundational work for this adaptation is described in Appendix B.

This work is currently focusing on the physical power system but can be extended to the cyber domain as well in future work (e.g., network triggering conditions and corrective actions such as rerouting packets, restoring default device configuration).

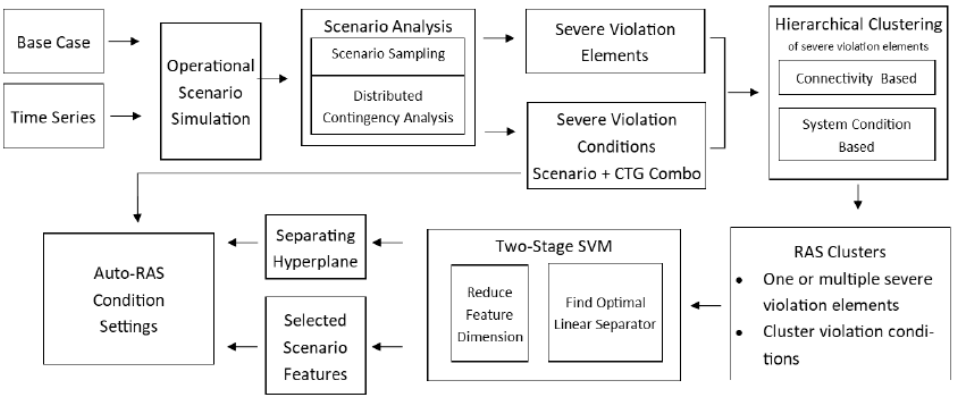


Figure 14: Diagram of autoRAS condition setting process.

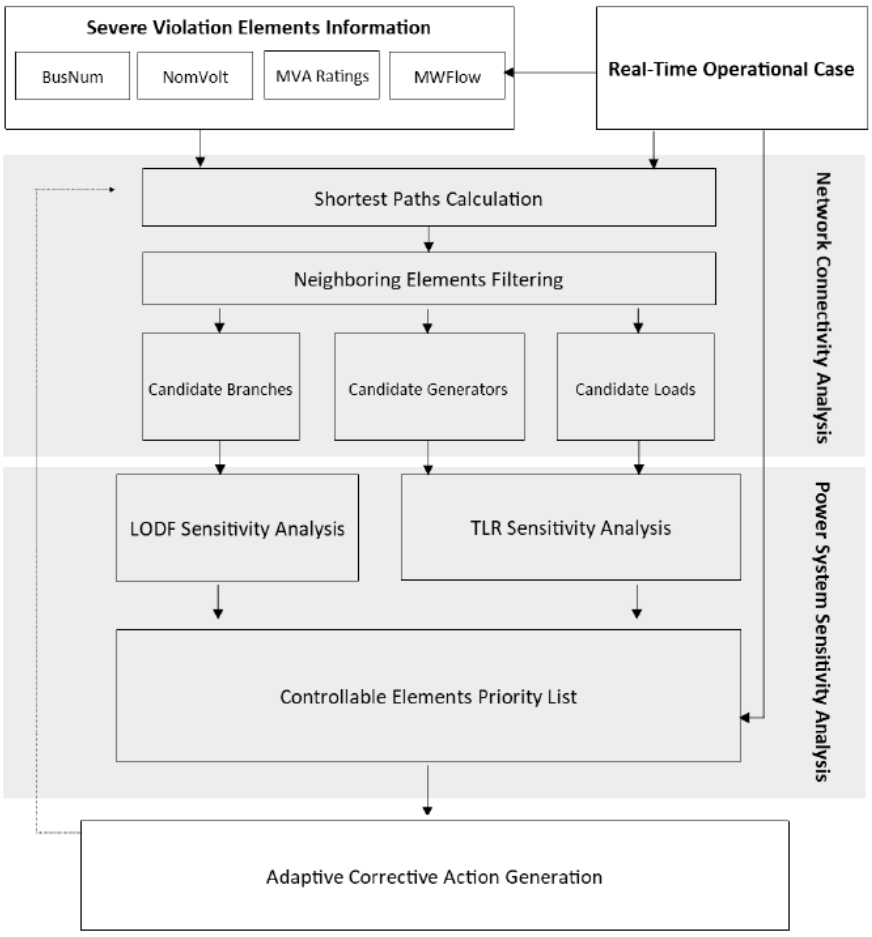


Figure 15: Diagram of autoRAS corrective action creation process.

5.3. Consensus Algorithm-Based Relay Voting (CARV) Scheme

To achieve selectivity and security in the HARMONIE-SPS approach, we developed a next-generation relay voting scheme leveraging consensus algorithms. The consensus algorithm-based relay voting (CARV) scheme extends traditional 2-out-of-3 voting schemes to a distributed system and aims to achieve consensus on system state and voting on response actions.

With the integral role of protection schemes in grid reliability, it is important for relays to operate correctly when they do take action (e.g., opening/closing breakers). Voting schemes can be used to compare trip decisions from different relays, for the same measurements, and to apply logic, such as two-out-of-three, to determine the final trip decision. In this manner, confidence in the trip action can be achieved and redundancy is obtained with the usage of multiple relays. The relays can be connected in series or parallel, depending on the logic used for the final trip decision [49]. This type of relay implementation is most commonly seen in transmission systems, where redundancy for reliability purposes is paramount. The placement, amount of redundancy, and addition of advanced features, such as communications-enabled relays, can vary depending on deployment within transmission systems versus distribution systems versus microgrids.

One method commonly utilized to improve the reliability of important components in a system is to use voting schemes. In such a configuration, the device in question is duplicated and several identical copies utilized to perform the same operations. This comes in several configurations, such as 2 out of 3 (2oo3), 2 out of 4 (2oo4), or other varieties. The output of each individual component is then wired together to vote on the outcome. This setup improves the reliability of the system overall by ensuring that the failure of a single component does not affect the performance of the system overall, and so is used in a variety of applications that require fault tolerant systems.

In more concrete terms, the probability of failure for the system is reduced by redundant connections. For instance, for the 2oo3 configuration system failure requires 2 components to fail in order to have the system fail as a whole. In this instance, these components are setup as parallel As an example, if the failure probability of a component at certain time is 0.02, then the probability of system failure is now 0.0004 since it requires 2 components to fail simultaneously. This is highly dependent on system configuration and assumes failure events would be mutually independent, which is not always the case.

However, for a cyber-physical electric grid where resilience is key due to the rise of cyber attacks and cascading impact of disturbances such as extreme weather and equipment failure from grid-edge systems, it is important to have an adaptive and resilient voting scheme. Additionally, we are interested in transmission system protective relays for use within an adaptive, online real-time SPS for automated response.

To develop a next-generation relay voting scheme, we leveraged consensus algorithms to incorporate inter-relay relationships and out-of-band data to better protect the power system as a whole [50]. Distributed calculation of system values, such as with distributed averaging, can be utilized to provide information that relays can check each other's values and ensure that overall an entire group of relays is reaching agreement on the state of the power system. Furthermore, by incorporating knowledge about byzantine faults in distributed computing, we can develop new voting scheme algorithms for relays to use to agree on protective actions to take following various

protection schemes, such as under frequency load shedding. By adding in this type of design into the system, relays are able to check and verify each other's actions, and alert when other relays in the voting scheme fail to operate correctly. This allows for faster detection of mitigation of failures and potential security issues.

5.3.1. CARV Design

An important design parameter for the CARV scheme described here is the configuration of the relay voting groups. In large bulk power systems, scalability is important for this scheme to be able to be broken down into smaller pieces that work together to implement the main portions of reaching consensus on system state and on remedial action conditions. These groups of relays may still pass information across groups, but at a much lower rate. This helps to ensure that the communications overhead from the relay voting is minimized and remains feasible.

In order to determine how the relay groups are configured and which relays belong in which group, as well as the ideal size for each group, there are multiple considerations to take. The first consideration has to do with the resilience of the overall system and how many relays can be permitted to fail without affecting the overall result of the consensus algorithm. For a desired number of failures f to be resilient to, the minimum voting group size is $n = 2f + 1$. This ensures that even with f relay byzantine failures, a majority of the relays will still be able to reach agreement on the correct solution for the group result overall.

This does not yet include other concerns for the group configurations, such as available infrastructure and network configuration, requirements for an individual RAS scheme, including but not limited to minimum latency requirements for the intragroup communications and relay locations for control of the remedial actions to be taken. Many of these requirements are dependent on the individual RAS scheme being implemented and the characteristics of the underlying power system, so we will briefly describe these considerations in our use case in this report and will present results on the communications overhead from implementing this relay voting scheme but will leave more complete analysis of timing considerations for voting group configurations to later work.

Note that even though the bulk of the relay communications required by this scheme is within each voting group, inter-group communications might sometimes be required. This will generally involve each voting group within a larger system sharing the group consensus results with other relay groups in order that the entire system can pass information concerning the RAS to all the relays in the system, even if a voting group does not contain members which are in the appropriate locations to measure certain system state values. This step is separate from voting within the group, but it remains important to ensure that it does not introduce a common mode failure point for the relay voting scheme at large. This can be done using random communications between groups for information about the global system state that needs to be shared and can be implemented by having a relay either query a random relay in another voting group or setup to have relays broadcast their current values to be shared to a random relay. Either of these implementations results in a random graph between voting groups, which can be shown to be convergent to consensus as long as each node is strongly connected

There is a trade-off here between speed and reliability of the voting system versus the amount of communications required. By including the ability to make voting groups subsets of the entire system, this algorithm can be tailored for each RAS implemented in the power system and to the existing infrastructure. Furthermore, this type of voting algorithm can be scaled for larger systems such as would be seen in real world applications.

Furthermore, a timeout functionality is included to ensure that each individual relay does not perform another inter-group query until others have had a round, ensuring that a compromised relay cannot do multiple fake reports in quick succession in a group without being flagged for non-compliance by the other relays in the voting group. This ensures that the data being reported will be accurate as long as a majority of the relays are working, which is a pre-existing requirement and so the resilience of the system is not reduced by these inter-group communications.

5.3.2. Distributed Calculation

An important step to implementat any RAS is to ensure that required up-to-date information about the system state for arming and triggering conditions is passed to the right location. If this is not done correctly, or the information cannot be trusted, then the RAS may fail unexpectedly. For this reason, a step is included in this relay voting scheme for the distributed calculation of important system values, including options for both system parameters and for measured system state values.

This is done in each relay voting group by having the relays work to reach agreement using matrix-weighted consensus algorithms for distributed calculation of system measurements, parameters, or other values of interest. This allows the relays in the system to share knowledge and check that values being reported are reasonable. If an individual relay is reporting values well outside the consensus result, there is sometimes the possibility of that being flagged at this step in the algorithm. That is however highly dependent on the characteristics of the underlying power system and location dependent dynamics and considerations. For instance, in an exemplar use-case, the variable used for the RAS is system load in two control areas, which can only be measured by calculating load at each individual bus in the system. These values are highly dependent on location, with only a few buses actually connected to loads, so there is a high variability in the values reported by each relay and detecting relay failure from discrepancies in reported values is not possible. In other cases, such as if we were utilizing power flow or voltage measurements, such an analysis for detection of anomalous behavior would be possible and can be included.

To compute a consensus value for the relay voting group, algorithms such as distributed averaging can be utilized. There are two variations of interest, for averaging of initial values, such as for system parameters, or for computing running averages for dynamic state variables that are measured and change over time. These two versions of distributed averaging can be written as:

$$x[k+1] = Ax[k] \quad \text{(Eq. 1)}$$

$$x[k+1] = A(p\,x[k] + (1-p)y[k]) \quad \text{(Eq. 2)}$$

where $y[k]$ is the sensor values at time k, and $x[k]$ are the consensus values for each relay, the matrix A is the adjacency matrix for the consensus algorithm and p is an averaging factor. In the case of averaging a set of values that is set initially, only Eq. 1 is needed. For instances where new measurements are included periodically, Eq. 2 will be needed instead. In this work, we will utilize Eq. 2 and will examine the choices for A and p when discussing use-case results.

A more complete description of utilizing these two variations of distributed averaging within this relay voting scheme is included in the initial development of the algorithm in [50], as well as analysis of results and their differences. It will depend on the needs of a specific RAS which consensus algorithm is required, or if another distributed calculation is needed. Note that if another type of calculation besides averaging is required for a RAS implementation, it is important that the calculation of values be performed in a distributed manner with each relay working independently or else the resilience benefits of utilizing this relay voting scheme will be compromised.

To minimize communications overhead, relays are split into groups with a distinct separation in communications within the group and between the groups. Intra-group communications are broadcast, while inter-group communications are configured to be sent to random relays in other groups. This can be thought of as examples of fixed network consensus algorithms for communications within a voting group, and random network consensus for inter-group communications. Further details and requirements for this formulation can be found in our full, submitted journal paper [51].

5.3.3. Relay Voting

The second portion of the scheme is the voting scheme itself, which ensures that each relay is not only able to vote on whether conditions warrant remedial actions but will know the result of the voting. Because of this, failures of relays to communicate with their peers or to implement remedial actions can be detected by those peers and alerts created that something is wrong. This allows not only for the distributed detection of relay failures but also creates opportunities for secondary remedial actions to be implemented. With the ability to take remedial actions even if individual relays fail, the resilience of the power system is enhanced and it becomes more difficult to create hazardous conditions. Note that a central part of this paradigm is that the relays are operating independently and there is no central point of failure, but they are engineered to cooperate to meet the system protection goals.

This is incorporated by utilizing the Practical BFT algorithm (PBFT), developed in [52], and the Robust BFT (RBFT) algorithm, developed in [53], to define the process by which relays communicate to their peers on whether a RAS should be armed and/or triggered. Note that the specific algorithm steps used in our approach generally follow the format as laid out in PBFT but is slightly modified to better match the requirements of a relay voting scheme with requirements based on which relays can take part in mitigations. Since there are specific limitations based on the locations of individual relays, the constraints of the power system topology need to be built into the overall approach. Similarly, it is important to ensure that no individual relay can become a common point of failure, which is why we also draw from the RBFT variant of the Byzantine Fault Tolerant algorithm.

There are no clients in our approach, as instead the relays create requests for voting rounds based on local conditions and protection scheme requirements, or at predetermined intervals. This "primary" relay(s) in this configuration is the relay that is responsible for performing the remedial action, while secondary and tertiary relays can be defined as well for relays that can perform additional mitigations if needed.

This entire process can be seen in Algorithm 1 pictured in Figure 16.

Algorithm 1 Relay voting with BFT

1) Relay i detects under frequency conditions
2) Relay i initiates request
3) Request for voting multicast to all other relays
4) All relays compute protection scheme calculations, determine load to shed
5) Each relay multicasts result to all other relays in group
6) Each relay waits for $f + 1$ replies, saves result.
7) Relay j that needs to shed load acts accordingly

Figure 16: Relay voting process with BFT.

The primary relay will wait until it gets $f + 1$ replies, where f is the max number of allowable node failures for a correct response from a voting round. In other words, this relay waits for a majority of relays in the voting group to send their outcome results and will utilize the majority answer. Any discrepancies can be flagged and reported as an alarm or indicator that a relay requires maintenance or could be compromised. Similarly, if a relay does not respond in a timely manner to a voting round, it could be flagged as unresponsive.

To ensure that individual relays are not able to cause the failure of the entire RAS, secondary and tertiary protective actions can be programmed into the system. Similarly, PBFT suffers from possible failure if the primary node in the system misbehaves and does not order voting rounds. To mitigate this issue, we borrow ideas from RBFT and do not specific a primary relay for voting rounds in the algorithm. Instead, we include two separate mechanisms for relays to vote on whether to arm or trigger a remedial action. First, relays will have voting rounds scheduled periodically at regular intervals, but also any relay can initiate a voting round as well.

Instead of utilizing primary nodes for initiating requests for voting rounds, the CARV algorithm splits relays into primary, secondary, and tertiary relays based on their location and their role in the RAS. This allows the RAS to include additional mitigations beyond the main responses in case important relays fail or do not react correctly. Note, this listing of relays does not affect how the relays communicate in the voting rounds at all and is purely an aspect of the RAS design and the power system infrastructure.

If the relays in a voting group detect that a relay is not responding or is not calculating correct values for the consensus algorithm, that relay will be flagged by the other relays in the group. This may include sending alerts to a control center and moving to secondary or tertiary mitigations if that relay is involved in the primary RAS response. One mechanism that the relays in the consensus algorithm utilize for detection is thresholding of the consensus values, as described further in [51]. The results in Section 7.3 will show how this impacts the consensus algorithm using the WSCC 9-bus system use-case. The overall CARV process is summarized in Figure 17.

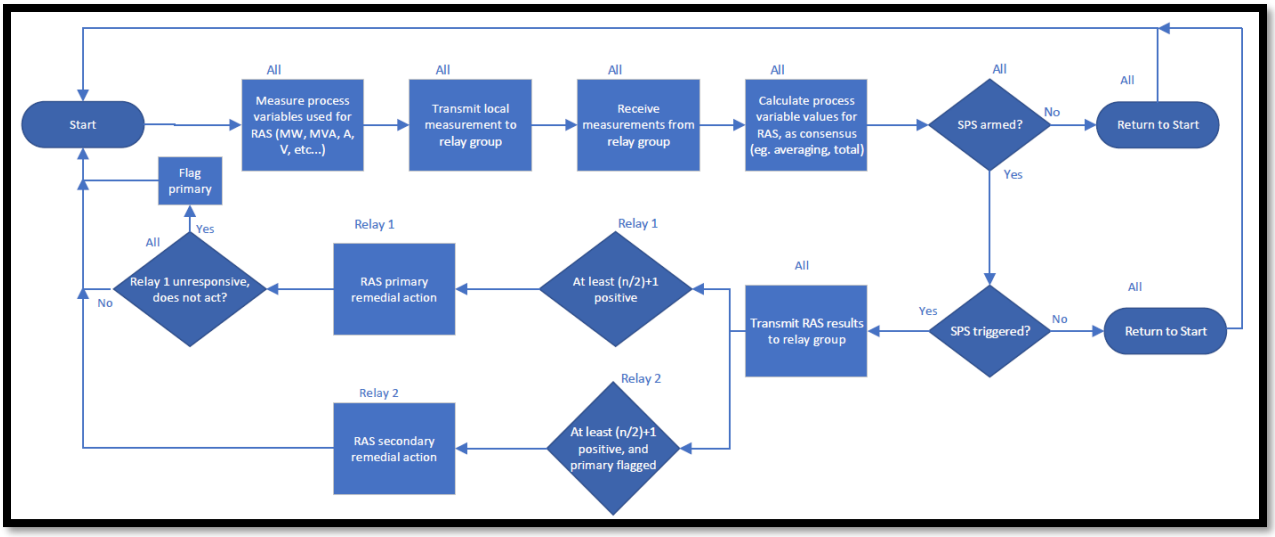


Figure 17: Overall CARV process.

5.4. Adaptive Routing Optimization Algorithm

To mitigate the undetectable threats, prevent the attack propagation, and improve the resilience of the system operation, an adaptive routing algorithm is developed to actively optimize the routing path of the communication system, based on the feedback from a state estimator (SE). The algorithm is leveraging the inter-substation-level cyber topology, along with the SE residues, to intelligently adjust the settings of controllable network components, i.e., routers and SDN switches, to reroute the traffic in real-time. The residues from the SE are first clustered based on the source, and then normalized and assigned as edge weights according to the original directed routing graph. A modified network simplex algorithm is developed to not only takes the edge weights (normalized residues and latency) into account, but also consider the bandwidth limit for each link.

To validate the algorithm, we initially tested in the WSCC 9-bus system with single-point false data injection attack, and then upscaled to the IEEE 39-bus system with single-point and multi-point attacks. By comparing the average state estimation residues between applying the proposed algorithm and the conventional shortest distance path (SDP) based schemes, it is very clear in Figure 18 that the proposed method could effectively isolate the threat and stop the harmful lateral movement at very early stage, as long as it caused any observable fluctuations on the SE residues. These results are detailed in Section 7.2 with the FDI attack scenario.

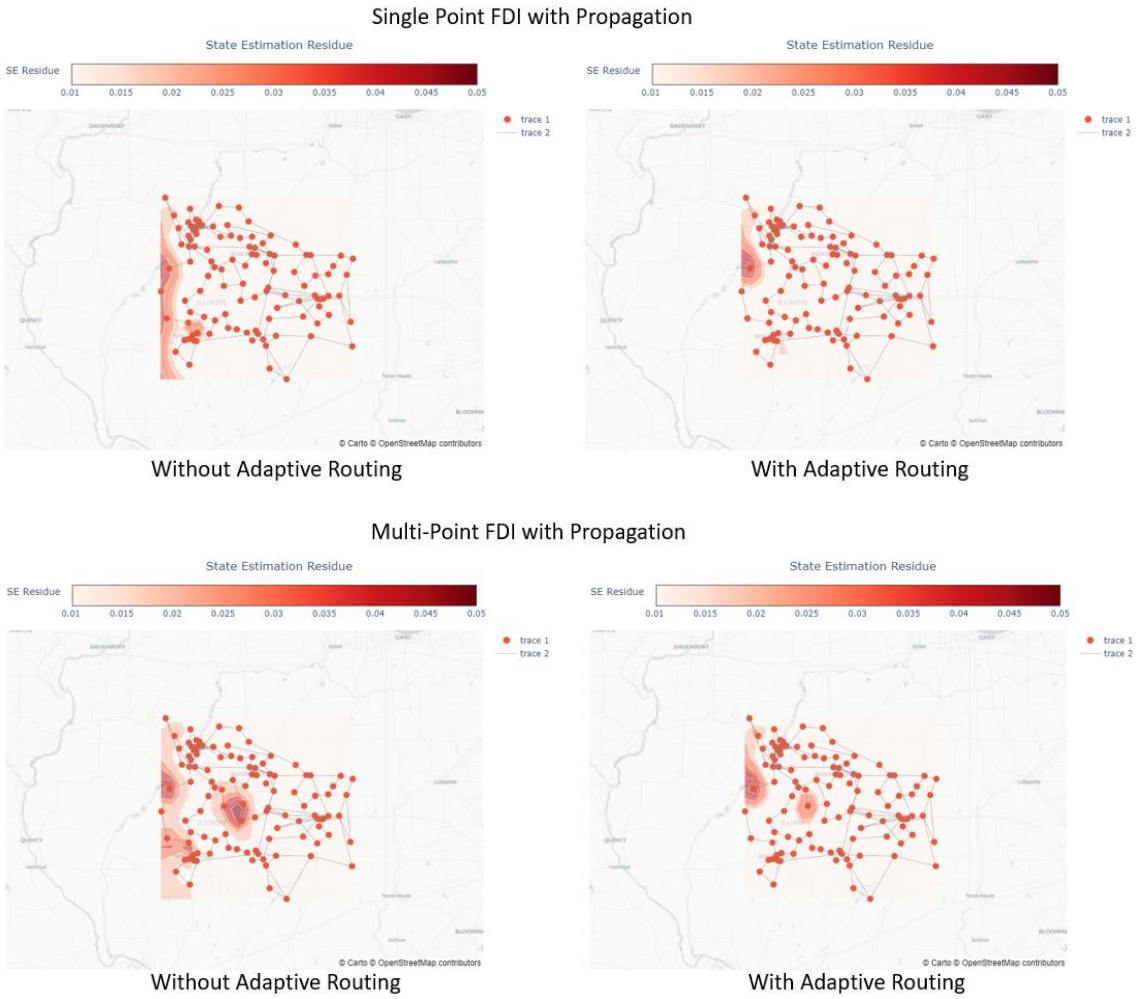


Figure 18: SE residues with and without adaptive routing for single-point and multi-point FDI.

6. TRAINING AND TESTING IN EMULATION

To test the HARMONIE-SPS approach, include the classification of different system disturbances and deployment of corrective actions, it is important to have an interactive environment where we can model various disturbances and collect and monitor real-time cyber-physical data streams.

6.1. Texas A&M University RESLab Environment

The TAMU RESLab testbed, visualized in Figure 19, is a high-fidelity interactive cyber-physical co-simulation platform with hardware-in-the-loop capability [54], [55]. It is leveraging Common Open Research Emulator (CORE) for network emulation, PowerWorld Dynamic Studio for electromechanical transient simulation, and Hierarchical Engine for Large-scale Infrastructure Co-Simulation (HELICS) for real-time data synchronization. On top of its federated simulation infrastructure simulating the inter-coupled cyber and physical systems, a control system layer also co-exists, to support the operation of the simulated grid. A SCADA prototype system, including RTU and PMUs, and EMS functions, like state estimation, automatic generation control, contingency analysis, voltage stability analysis, etc., are integrated in the control system layer. The whole testbed is interactive and could be running in the real-time mode. Physical devices (generators, transformers, loads, etc.) and cyber devices (routers, switches, firewalls, etc.) can be adjusted on-the-fly.

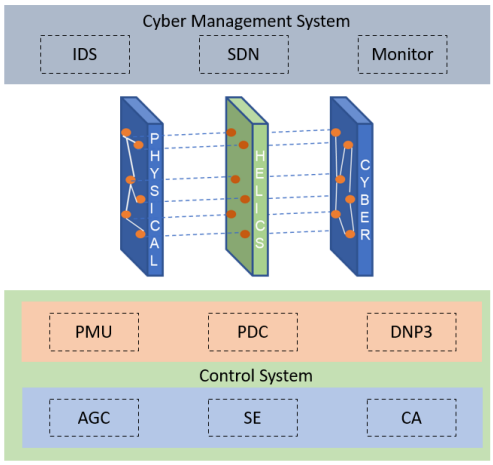


Figure 19: RESLab cyber-physical testbed overview.

The TAMU testbed has been used to provide physical measurements and network data from various cyber-physical scenarios (natural (e.g., extreme weather) contingencies, cyber contingencies, and cyber-physical events) for the HARMONIE project. Analog phasor measurements, including voltages and currents, along with network packets and traffic data, were captured in the testbed and then passing to the detection algorithm for learning and inference purposes. Several cyber and cyber-physical contingencies, including denial of service attack, time delay attack, false command injection attack, false data injection attack were simulated in the testbed to provide credible and reproducible data for the HARMONIE-SPS project.

6.2. Synthetic Cyber Topology Generation Tool

When developing cyber-physical emulation environments, a challenge we commonly face is the lack of benchmark cyber system topologies. For example, there exists various IEEE benchmark cases for

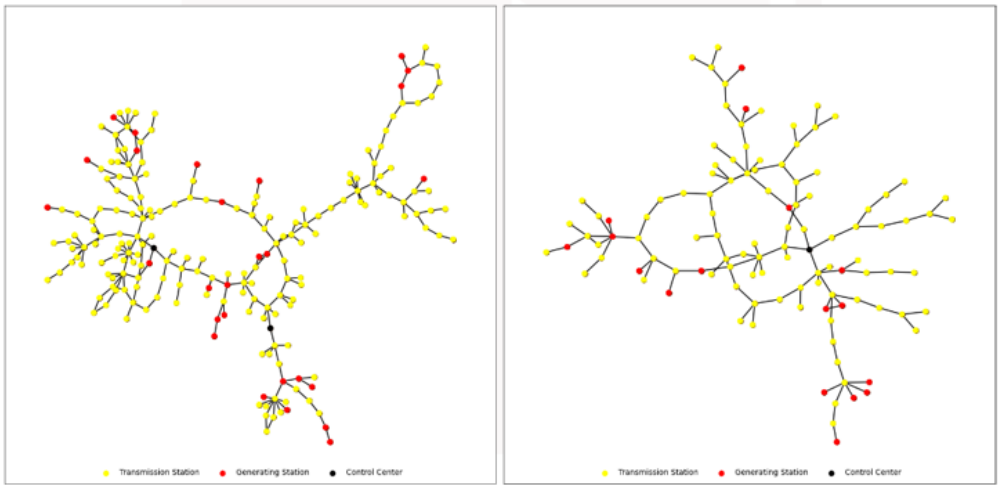
power systems that are used across industry and academia. However, those power system models do not have associated communication network designs. We needed to develop realistic cyber topologies to pair with power system benchmark cases (e.g., WSCC 9-bus, IEEE 39-bus).

Thus, we developed synthetic inter-substation cyber networks that match real industrial cyber networks with respect to cyber-physical intercoupling, graph characteristics, and network properties. A three-stage algorithm was proposed to effectively and efficiently construct a realistic underlying cyber network for any given transmission network models, leveraging mixed-integer programming, heuristics exploration, and graph theory. The first stage is to generate a sequence of degree which not only has the prescribed degree distribution, but also satisfy all the constraints of a connected graph [56]. For instance, the Havel-Hakimi inequality constraint needs to be satisfied to ensure that the sequence is capable to generate a simple and connected graph. The second stage is to generate the connected graph using the given degree sequence. Configuration model is used to generate a simple graph that matches the degree sequence, then edge swap technique is utilized to connect each isolated subgraphs without changing the node degree. An epsilon-greedy algorithm is developed to shuffle the edges so that the graph properties can be iteratively getting closer to the desired values. The pseudo code is shown as below:

Algorithm 1: Shuffle the edges via ϵ-greedy exploration

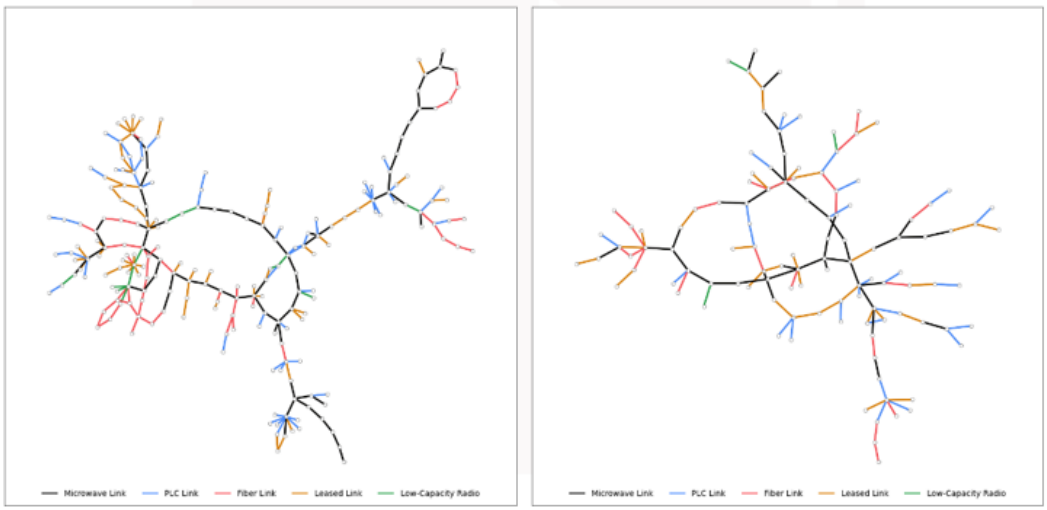
```
Q, MaxIteration, EarlyStopCriteria, ϵ
while it < MaxIteration do
    n ← uniform random number between 0 and 1
    ϵ ← setting new epsilon with ϵ-decay
    if n < ϵ then
        A ← random edge swap on the base graph
    else
        A ← random edge swap on the graph which has max(Q)
    end
    if A has self-loops or islands then
        Continue
    else
        Q ← rewards based on the distance to the prescribed graph metrics, current graph
    end
end
```

The last stage is to assign the different node types (substation and generation station) and the edge types (microwave, PLC, fiber, leased and low-capacity radio) by leveraging the mixed integer programming, so that the generated network is tightly close to the real network from both homogeneous and heterogeneous graph perspectives, as shown in Figure 20 and Figure 21. The resulting cyber topologies for the WSCC 9-bus and IEEE 39-bus systems are provided in Section 6.3 and 7.2.



Reference model Generated model

Figure 20: Transition from reference to generated model based on homogeneous topological properties.



Reference model Generated model

Figure 21: Transition from reference to generated model based on heterogeneous tolological properties.

6.3. RTDS and SCEPTRE™ Emulation Environment

A high-fidelity cyber-physical emulation environment was constructed using SCEPTRE™ and a real-time digital simulator RTDS. SCEPTRE™ is a Sandia technology that enables implementation of different ICS communication protocols such as DNP3 and Modbus [18]. We are interested in collecting data from the power system at different sampling rates due to non-contingency (low-sampling) and contingency events (high-sampling). The RTDS is able to stream *C37.118* data that is collected in virtual phasor data concentrator (PDC) database; this database is then tapped into by SCEPTRE™ to update changes to the communication network and ICS devices [19]. The representative communication network was developed using the synthetic cyber network tool described earlier.

The emulation enables modeling and testing of cyber-physical disturbances and the ability to extract high-fidelity data from both the cyber and physical systems. Furthermore, this data can be used to train and test the machine learning framework's ability to classify system conditions and deploy suitable cyber-physical corrective actions. We also incorporated hardware-in-the-loop (HIL) equipment such as digital relays (i.e., SEL 451, SEL 421) into the environment and tested the CARV scheme. Additional details about the emulation construction and initial testing can be found in [57].

Example use-cases and associated environment architecture are presented next. Figure 22 shows an example emulation architecture with the WSCC 9-bus system and a SEL-451 relay as HIL. Figure 23 shows the emulation networking diagram associated with the WSCC 9-bus system. The overall cyber-physical system mapping is shown in Figure 24 with a cyber topology generated with the synthetic tool discussed earlier in the report.

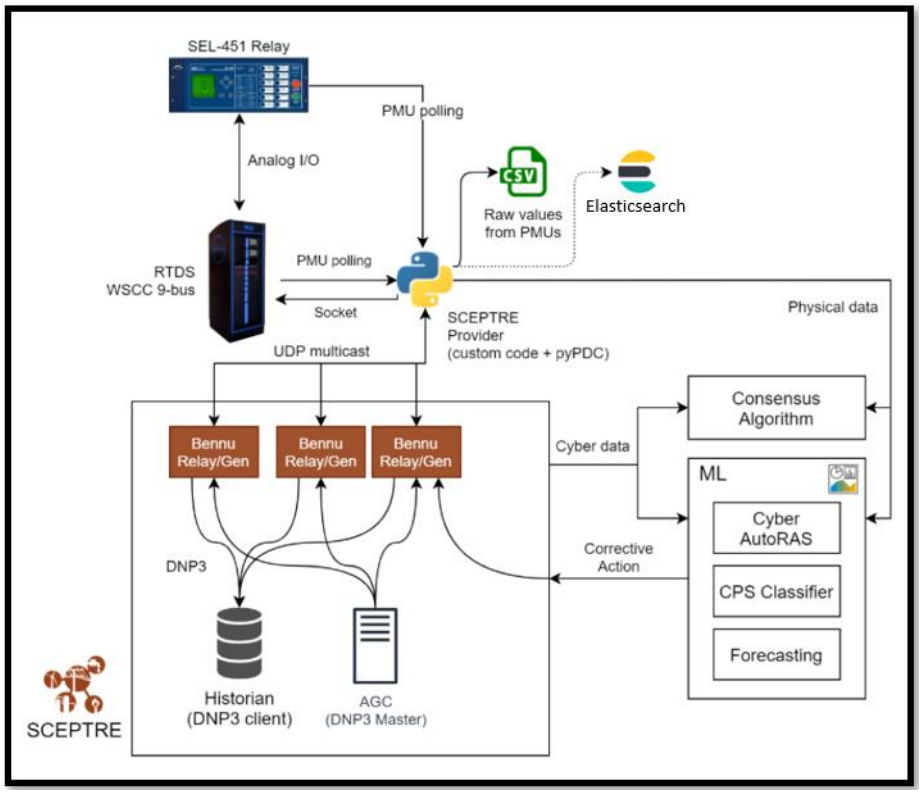


Figure 22: Exemplar emulation architecture for WSCC 9-bus system.

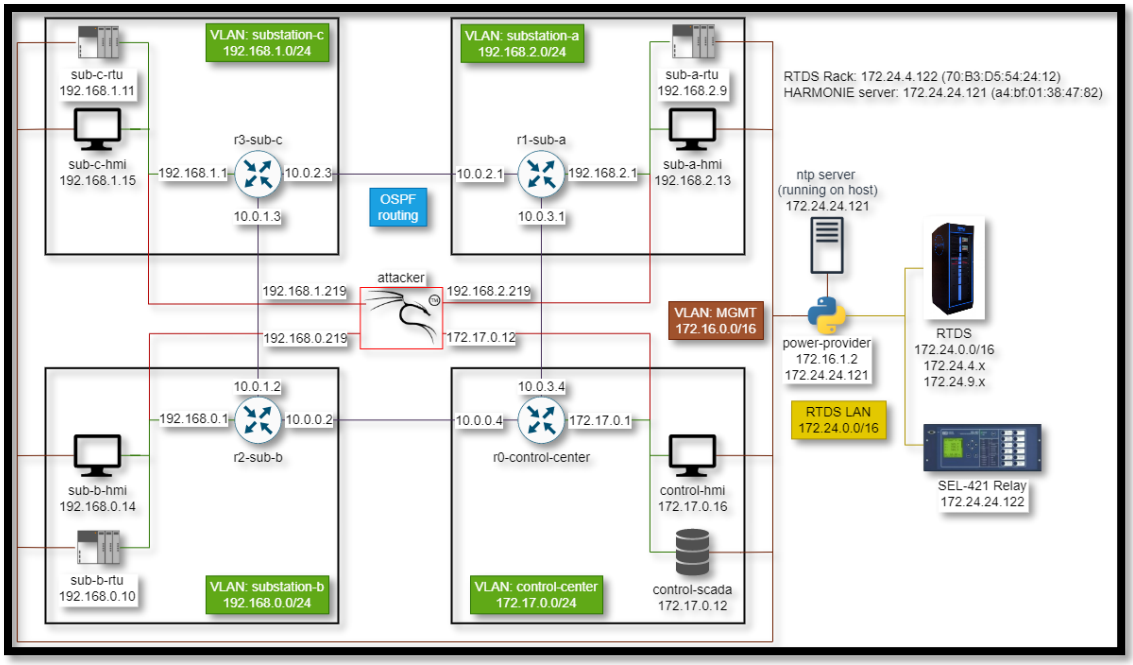


Figure 23: Networking diagram of WSCC 9-bus system.

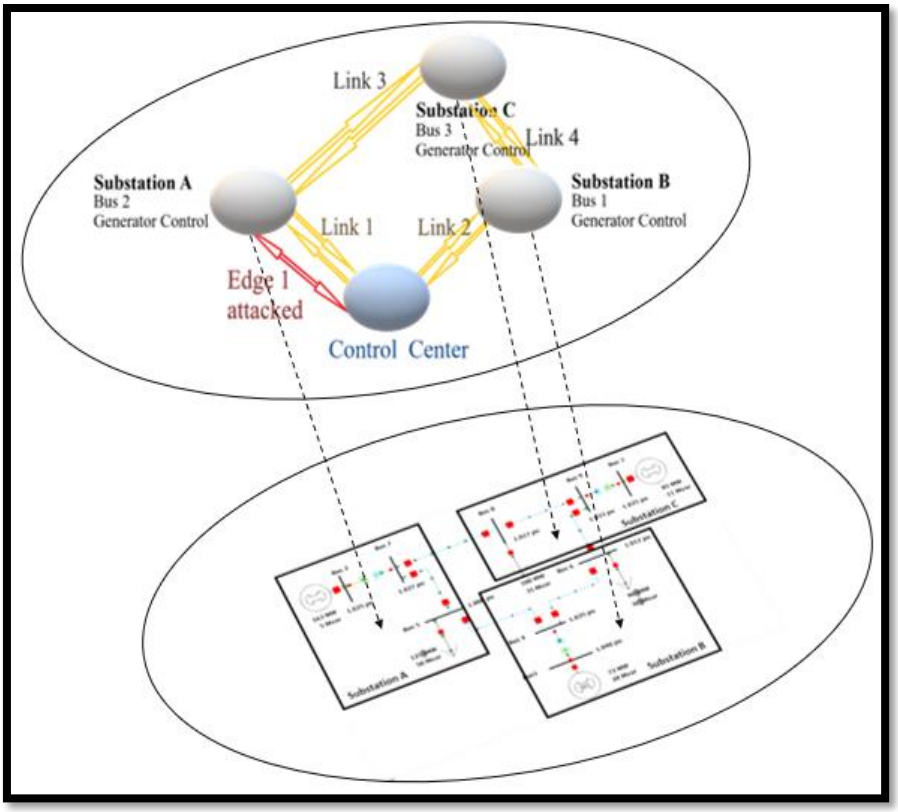


Figure 24: Overall mapping of WSCC 9-bus system cyber-physical emulation environment.

6.3.1. Emulation Tools, Databases, and Visualizations

The RTDS and SCEPTRE™ emulation environment utilized a few different tools to support experimentation, visualization, and data collection. The next few subsections describe these tools and their uses.

6.3.1.1. ADROC

ADROC (ADvancing Resilience Of Control systems) is a LDRD project (PI: J. Thorpe) that is developing a cyber threat modeling and resilience experimentation platform for characterization and quantification of cyber vulnerabilities and risks. The ADROC platform uses multiple levels of fidelity for system and threat modeling in order to achieve a more efficient analysis (both cyber and physical). Low-fidelity math models are used to down-select scenarios to those of higher interest, and high-fidelity emulation is applied to model these high-interest scenarios in a way that can be more deeply understood and quantified.

With ADROC's ability to orchestrate these different experiments, the HARMONIE-SPS team collaborated with the ADROC team to assess the CARV scheme and 1) investigate the effectiveness of different configurations of CARV vs. no CARV and 2) investigate the robustness of the "best" CARV configuration against DoS attack variations. The scenario flow is visualized in Figure 25 where injects are used to determine if CARV and/or CALDERA starts. CALDERA is a is a cybersecurity framework developed by MITRE provides automated adversary emulation capabilities as well as other automated security assessments [58]. A DoS attack scenario during the operation of a load shedding scheme was planned for evaluating CARV. Cyber (network) and physical (voltage) data was to be collected to assess deviations between baseline and attack scenario data streams; for the attack scenario data streams, metrics would be outputted for the different CARV configurations.

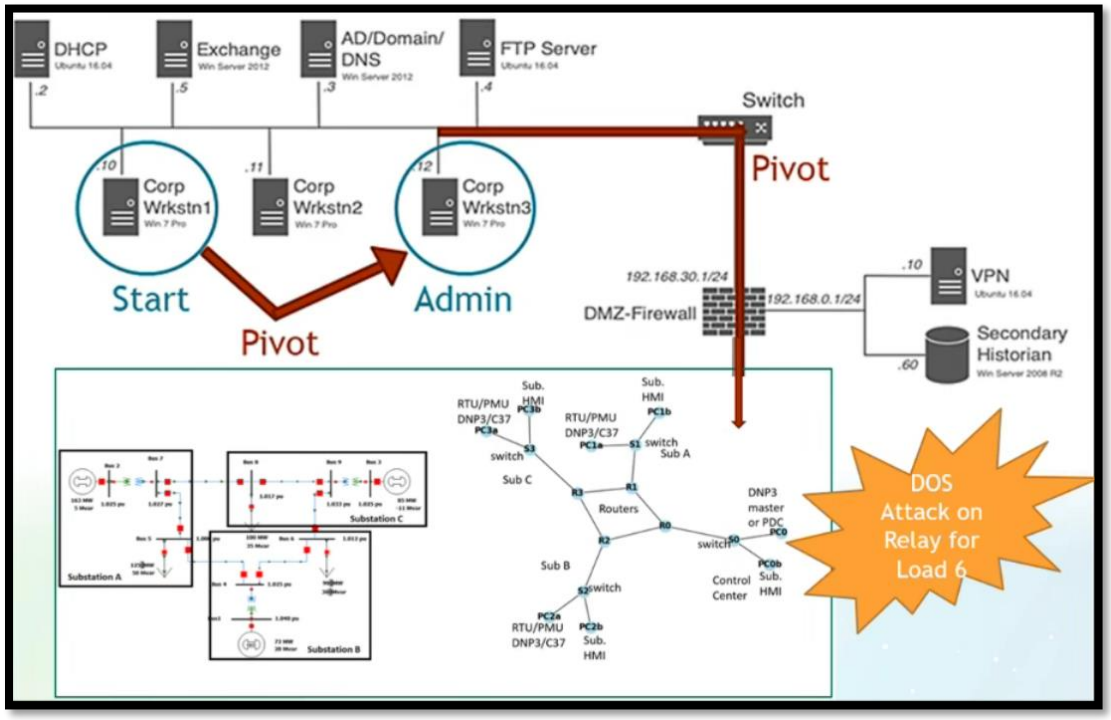


Figure 25: Scenario flow for ADROC and CARV experiments.

Due to time constraints, the ADROC and HARMONIE teams were not able to complete the experimentation but were able to integrate the ADROC topology with the WSCC 9-bus emulation built within the RTDS and SCEPTRE™ environment. This included the ability to collect cyber-physical data from the emulation, ensuring ADROC worked with an environment with a HIL relay (the SEL 421). A series of experiments were also designed to best evaluate CARV in the DoS attack scenario, which would have provided comprehensive metrics for its performance and informed the most effective, "best" configuration for this use-case.

All in all, although the experiments were not run, the design and integration work that was performed was very useful and will be continued in future iterations of ADROC and HARMONIE research. The mitigation-focused analysis using ADROC will not only aid HARMONIE research but related and new projects starting in FY23 (RES MC griDNA and EHS InterGraph-CPS LDRD projects); the teams are exploring avenues for this continued collaboration currently. Furthermore, this type of analysis can be extended to evaluate a suite of corrective actions, especially if multiple are deployed to combat a multi-hazard event (e.g., cyber attack during extreme weather event). The ability to extract metrics and compare baseline/attack data streams is essential for assessing adaptive mitigations. Lastly, this comparison can also be used as an out-of-band check for the HARMONIE-SPS machine learning framework and its ability to detect different abnormalities in cyber-physical data streams.

6.3.1.2. Elastic Stack Implementation

Elasticsearch is a search engine that supports distributed architecture and is developed alongside the data collection and log-parsing engine Logstash, analytics and visualization platform Kibana, and data shippers Beats; overall, as an integrated solution this is called the Elastic Stack [59]. In the HARMONIE-SPS RTDS and SCEPTRE™ emulation environment, the Elastic Stack was used to collect the cyber-physical data for baseline, attack scenario, and mitigation deployment; this data was then used by the machine learning framework to classify the system conditions. An example visualization from the environment using Kibana is shown in Figure 26.

Figure 26: Example Kibana visualization of cyber-physical data from emulation.

7. EMULATION EXPERIMENT RESULTS

The HARMONIE-SPS methodology was tested in a few different environments with two different use-cases throughout its development. The environments include Texas A&M University's RESLab and RTDS and SCEPTRE™ emulation as described earlier in the report. The use-cases are the WSCC 9-bus system and IEEE 39-bus system. The next few sections detail the results and insights learned from the experiments.

7.1. WSCC 9-Bus System Use-Case

To test the initial HARMONIE-SPS machine learning formulation, we utilized the TAMU RESLab environment with the WSCC 9-bus system and associated synthetic cyber topology, pictured in Figure 24. Additionally, ML results with the WSCC 9-bus modeled within our RTDS and SCEPTRE™ emulation are presented in Section 4. At a high level, the machine learning approach converts incoming data (cyber data or physical data) into a graph of interconnected nodes, where each edge is a flow of information with an associated timestamp. After the whole capture is split into subgraphs using 24-second sliding windows, the algorithm relies upon two deep learning architectures to obtain an overall representation of the system state in each window:

- A Graph Convolutional Neural Network (GNN), which applies deep learning to the structure of interconnected nodes in the subgraph, and
- A Recurrent Neural Network (RNN), which applies deep learning to the temporal ordering of the edges in the subgraph.

Using the GNN and RNN in tandem would theoretically allow for representing the changing system state over time with the RNN while understanding the connectivity of the cyber or physical network with the GNN, especially in complex networks. First, the GNN operates by passing four rounds of messages between edges, and the resulting edge vectors are passed to the RNN to encode temporal information.

To assess the classification of different system conditions within a cyber-physical grid system, we add a classification layer onto the network that predicts two binary labels: whether a cyber disturbance is occurring and whether a physical disturbance is occurring. This combination of two binary labels allows our model to categorize the system state into four categories:

1. Normal operations
2. Cyber-only disturbances
3. Physical-only disturbances
4. Cyber -physical disturbances

7.1.1. Initial Classification Results

Four different disturbance scenarios were modeled within the TAMU RESLab environment using the WSCC 9-bus use case. They included a mix of cyber-only, physical-only, and cyber-physical events.

1. Denial of service (cyber-only)
2. Single line-to-ground fault (physical-only)
3. Tripping command injection (cyber-physical)

4. Time-delay attack (cyber-physical)

A total of 50 network and physical data captures of various 2-minute scenarios of each of the listed disturbances were used to test the machine learning approach. For the experiments, we partitioned all scenarios into 30 for training, 10 for validation and model selection, and 10 for testing. These were then split into their respective sliding windows.

We ran experiments varying the size of the training data and comparing the results when using a model that has already been pretrained using some basic predefined perturbations versus a model that had not been pre-trained. Table 2 contains the results of these experiments for training with 900 windows (30 scenarios) and with 100 windows (rv3.3 scenarios). We used the area under the receiver operator curve (*AUC*) as our metric because it identifies how well a model's predictions split the two classes apart and does not require a predefined threshold to convert real-valued confidence scores into a discrete class prediction.

Table 2: Preliminary results for HARMONIE-SPS ML model (cyber anomaly AUC / physical anomaly AUC)

	With Pretraining	No Pretraining
900 windows	0.74 / 0.92	0.95 / 0.92
100 windows	0.49 / 0.64	0.52 / 0.60

From these results, we can see that using the full training data, our model can differentiate between disturbances and normal behavior. We also hypothesize that the pretraining step either adds nothing to or even mildly hinders the performance of the model, especially when identifying cyber anomalies. We attribute this to a domain shift between the inputs during the pretraining step, where perturbed graphs are given to the model, and the training step, where unmodified graphs are given to the model. That is, our current approach is closer to transfer learning than pretraining. We also provide a confusion matrix for the best model (using all 900 training windows with no pretraining) in Figure 27.

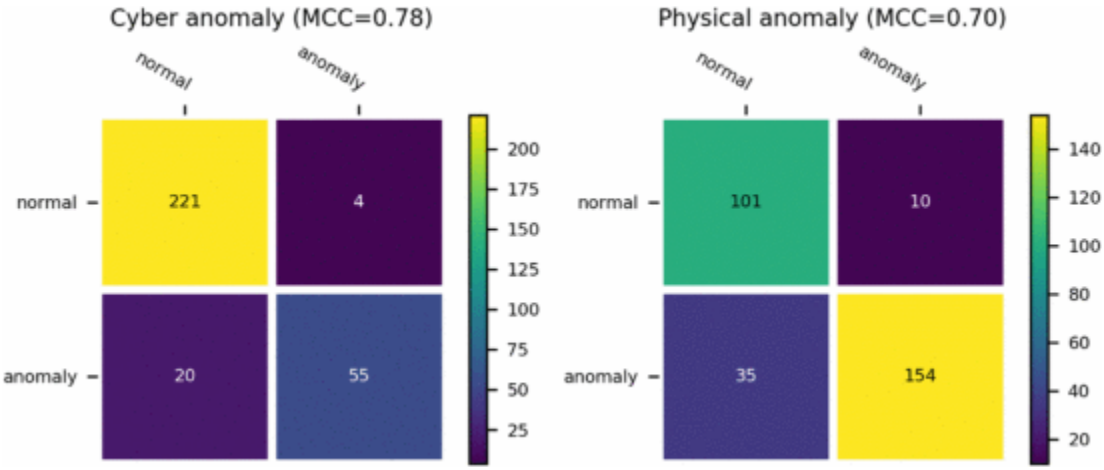


Figure 27: Confusion matrices for identifying cyber and physical disturbances on the test data using a threshold of 0.5. Matthew's correlation coefficient (MCC) is used to assess the quality of the predictions. Rows correspond to actual classes and columns correspond to predicted classes.

In Figure 28, we use our best model (trained on all 30 scenarios with no pretraining) to plot the predicted anomaly scores for each scenario in the test set. Since our approach uses 24-second sliding windows, all windows ending between 00:00:59 and 00:01:23 will contain the disturbance which occurs at 00:00:59 (blue vertical line). Note that some scenarios have cyber disturbances only in the middle of the capture, which is why some cyber anomaly scores drop to nearly 0 after 00:01:24. In these plots, we see that our deep learning approach can roughly identify when a disturbance occurs and whether the disturbance is in the cyber, physical, or a cyber-physical event. Further context and details of these results can be found in [60].

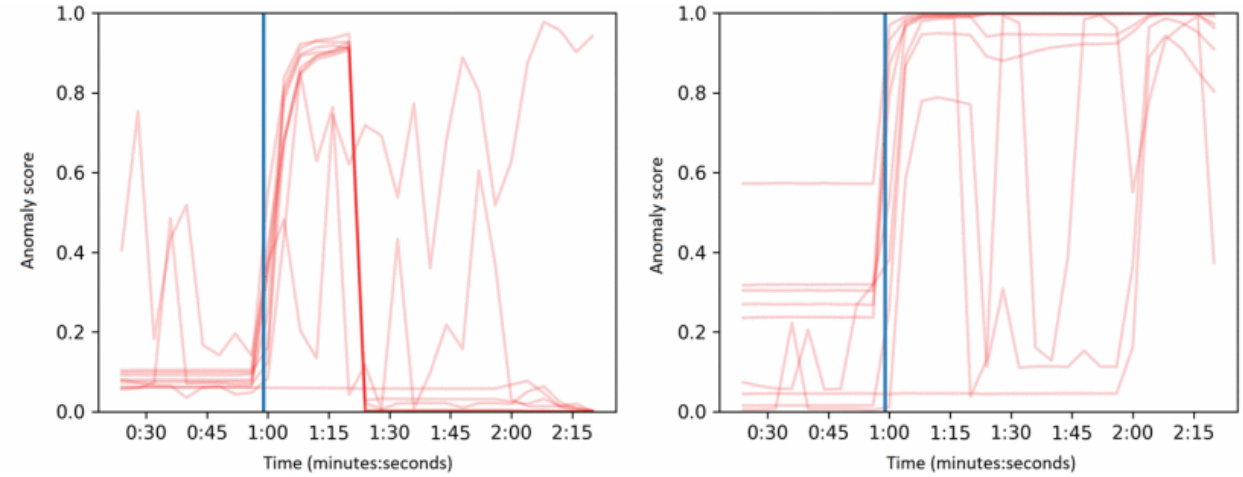


Figure 28: Reported anomaly scores over time for the 10 test scenarios. A value of 1 indicates confidence in an anomaly and a value of 0 indicates the confidence of normal operations. Left: cyber anomaly score. Right: physical anomaly score.

7.1.2. Graph Neural Network and Transformer Model Testing

After initial testing with the GNN and RNN, we also explored the inclusion of transformer models to resolve sequential and temporal differences in the cyber and physical datasets. Traditionally, a recurrent neural network such as a Long Short Term Memory (LSTM) [61] or Gated Recurrent Unit (GRU) [62]would be employed for neural processing of sequential data. In recent years, however, the Transformer model [21] has been shown to yield superior performance on most tasks, especially in the natural language processing domain, which deals primarily with long sequences of input. The Transformer architecture has also been successfully applied to physical systems [63].

We explored different implementation techniques of the Transformer architecture in [60]and performed several experiments to test the GNN and Transformer model efficacy for processing and classifying cyber-physical events. We leveraged the TAMU RESLab WSCC 9-bus use-case and tested DoS attack, false command injection (FCI) attacks, time delay (TD) attacks, and contingencies such as single-line-to-ground faults. Ultimately, we had 50 total scenarios, most of which contained cyber disturbances, physical disturbances, or both. Each scenario is roughly two minutes long with the disturbance (if present) happening at the one-minute mark. To allow our machine learning system to isolate rough temporal regions where disturbances happen, we treat each two minute capture as a training, validation, or test example and split it into over-lapping 30-second time windows.

To test the efficacy of our method, we trained 20 versions of the model on various slices of data. Only 50 scenarios were available to us, so to make the most of this small dataset we report all our

results using cross validation. We split our data into five random folds, each with ten scenarios. Since scenarios are further broken down into overlapping sliding windows, each window will not be independent from some others within the same scenario, so all sliding windows from the same scenario were placed into the same fold.

After this, we train independent models for each fold, withholding that ten-scenario test fold for evaluation. Within the four remaining training folds, we reserved one as the validation set for model selection, choosing the model which performed best on this validation fold. In summary, of the five folds, we assigned one as the test fold, one as the validation fold, and the remaining three as the training folds. In all of our experiments, one model was trained for each of these settings for a total of 20 distinct models.

To further reduce the high variance of our models incurred by training on such a small dataset, we elect to use bagging to combine multiple models into one. Specifically, we average the output of each of the four models trained on each test fold. The result is five aggregate models, one per test fold, each consisting of four models, one per validation fold. To reduce variance further, we ran each sliding window through the model four times, each time resampling which edges are kept in the Rationale Neural Network layer. The outputs predictions of all four runs are averaged to create the overall prediction for that sliding window. (Due to this innate randomness, these metrics are approximate and running the evaluation again on the same models would produce results differing by around 0.02.)

For each model architecture, we present Receiver Operator Curve (ROC) plots and Area Under the Curve (AUC) scores for detecting cyber disturbances and physical disturbances. Additionally, we include a confusion matrix for a decision threshold of 0.5 and its corresponding Matthew's Correlation Coefficient (MCC) scores. We also analyze the average percentage of edges that are selected by the Rationale Neural Network for propagation to the GNN and/or RNN. A low percentage indicates that the Rationale Neural Network significantly down-sampled the edges being used. A summary of the experimental results is included in Table 3 and in Figures 29-32. Detailed explanations and analyses of the experiments are provided in the full paper [30].

Table 3: Experimental results of each model

Architecture	Rationale %	Cyber Disturbance Detection		Physical Disturbance Detection	
		MCC	AUC	MCC	AUC
Traditional Transformer	39.3%	0.77	**0.98**	0.57	0.85
Random-windowed Transformer	46.1%	0.70	0.95	**0.63**	**0.87**
GNN	48.0%	**0.85**	0.96	0.18	0.68
GNN + Transformer	N/A	0.74	0.97	0.30	0.77

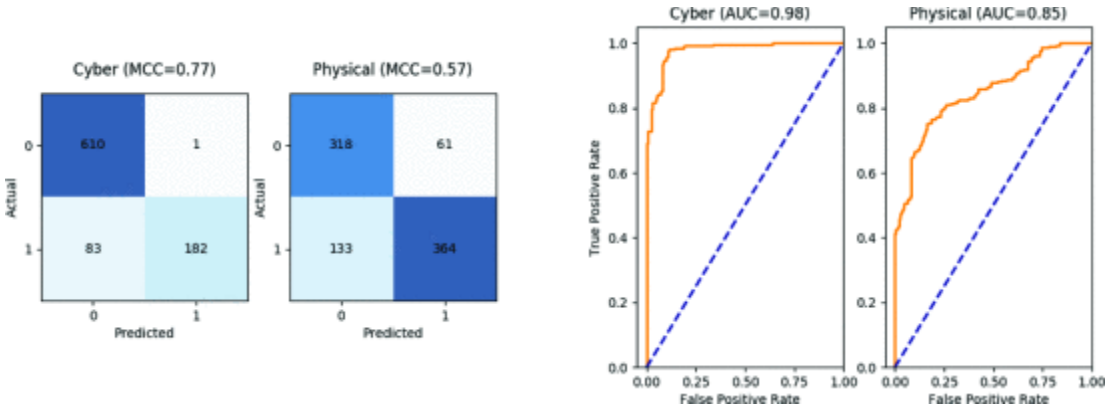


Figure 29: The confusion matrices, MCC scores, receiver operator curves (ROCs), and AUC scores for the traditional Transformer model detecting cyber and physical disturbances.

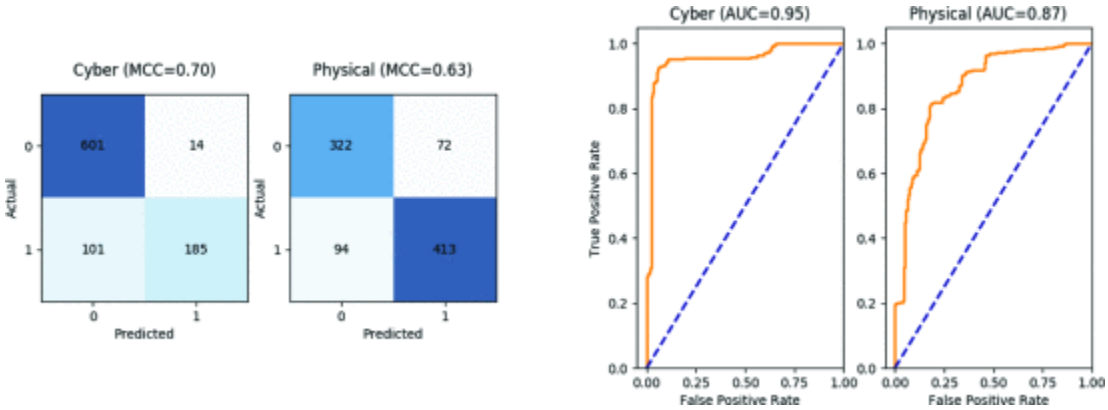


Figure 30: The confusion matrices, MCC scores, receiver operator curves (ROCs), and AUC scores for the random-windowed Transformer model detecting cyber and physical disturbances.

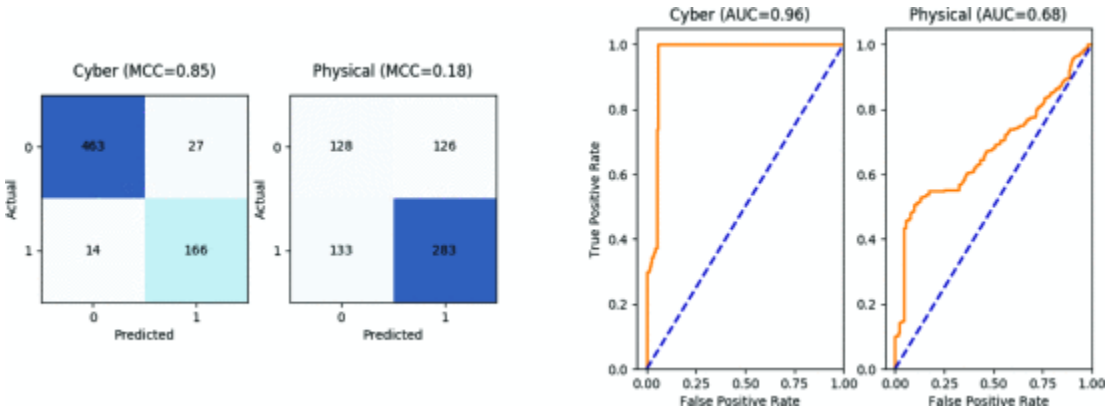


Figure 31: The confusion matrices, MCC scores, receiver operator curves (ROCs), and AUC scores for the graph neural network model detecting cyber and physical disturbances.

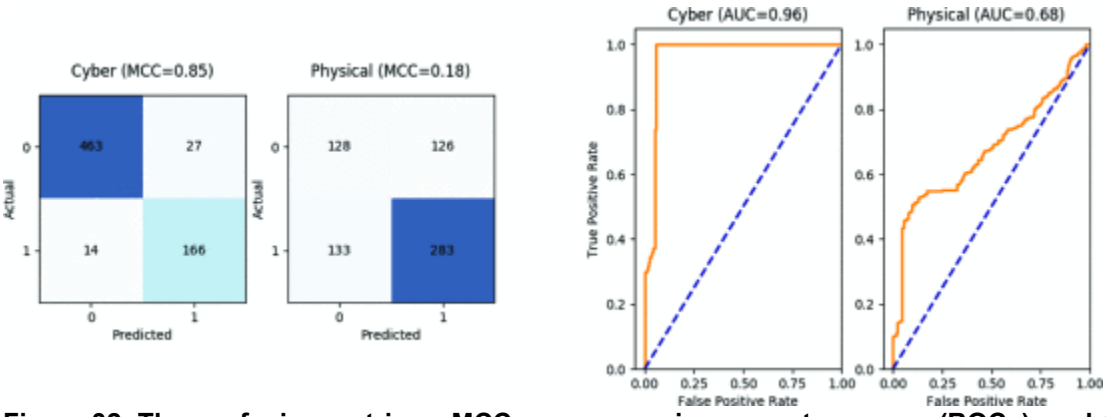


Figure 32: The confusion matrices, MCC scores, receiver operator curves (ROCs), and AUC scores for the GNN and random-windowed Transformer operating in series to detect cyber and physical disturbances.

These experimental results provide a number of insights that will guide the HARMONIE-SPS machine learning framework and can guide other SPSs in the future.

First, we see that in our scenarios and contingencies, cyber disturbances are generally easier to detect than physical disturbances. As noted above, we suspect that modeling the long-term trajectory of the physical data is a more difficult task for our deep learning model than searching for a small number of malicious or problematic network packets. Along these lines, as expected, we observe that the GNN has the most difficulty identifying a physical disturbance.

Second, we see that using the GNN and Transformer together in this way does not yield the performance increase we expected. While the GNN and Transformers perform well on detecting cyber disturbances and physical disturbances respectively, the GNN + Transformer model underperforms the best model in each of those categories. This suggests that while the GNN and Transformer model each contribute valuable information to the process, there is room to improve the way we link them together into a cohesive model.

Third, we see that the Rationale Neural Network kept 40-50% of the edges (packets or phasor datapoints). Upon closer inspection, the edges most often kept are TCP packets, and all phasor measurement edges seem to have been assigned approximately the same probabilities. While removing 50-60% of edges is a good start, we see value in exploring techniques to reduce the number of edges retained and thus further isolate the source of the disturbance.

In Section 4, we detail the final state of the ML framework for HARMONIE-SPS developed for this project where these testing insights were used to inform the latest implementation.

7.2. IEEE 39-Bus System Use-Case: False Data Injection Attack

A false data injection (FDI) scenario is developed to simulate a stealthy FDI attack on critical measurements in a power grid. The attack is designed to be stealthy, meaning it should not be detected by the conventional bad data detection (BDD) test. However, it will cause non-negligible variations on the state estimation results, which may further interrupt the normal operation of the grid. This type of stealthy attack can also indicate adversarial presence in the system and reconnaissance and capability testing activities.

A state estimator, fastSE, along with the corresponding BDD module, was developed to provide the state estimation results [64]. In order to be capable of running in a real-time environment, sparse matrix, memory pre-allocation, single instruction multiple data (SIMD), and improved pre-permutation are all applied to greatly improve the computational efficiency of the estimator. Figure 33 provides computational efficiency results before and after improvements to the state estimator implementation using a synthetic 200-bus system ('ACTIVSg200').

ACTIVSg200	dSbus_dV	dSf_dV	dI_dV	Constructing H	Compute dx
Before	47.9	39.8	37.8	74	70
After	3.8	5.8	5.7	23	50.2
Gain	24.2X	11X	9.8X	6.5X	1.5X

Figure 33: Computational efficiency gains with novel state estimator implementation.

The stealthy FDI scenario was implemented within the RTDS and SCEPTRE™ environment along with the fastSE state estimator, routing optimization algorithm to mitigate the FDI, and a few other additions to support the attack execution. Figure 34 summarizes the environment updates, including complimentary efforts such as the integration of CARV and ADROC into the environment. Additionally, to collect the FDI-compromised data streams, a "dirty" Elasticsearch database was implemented in the environment. In this manner, fastSE could process the "dirty" Elasticsearch data streams and detect the FDI attack but we could also collect the "clean" Elasticsearch database to compare the baseline and attack data streams.

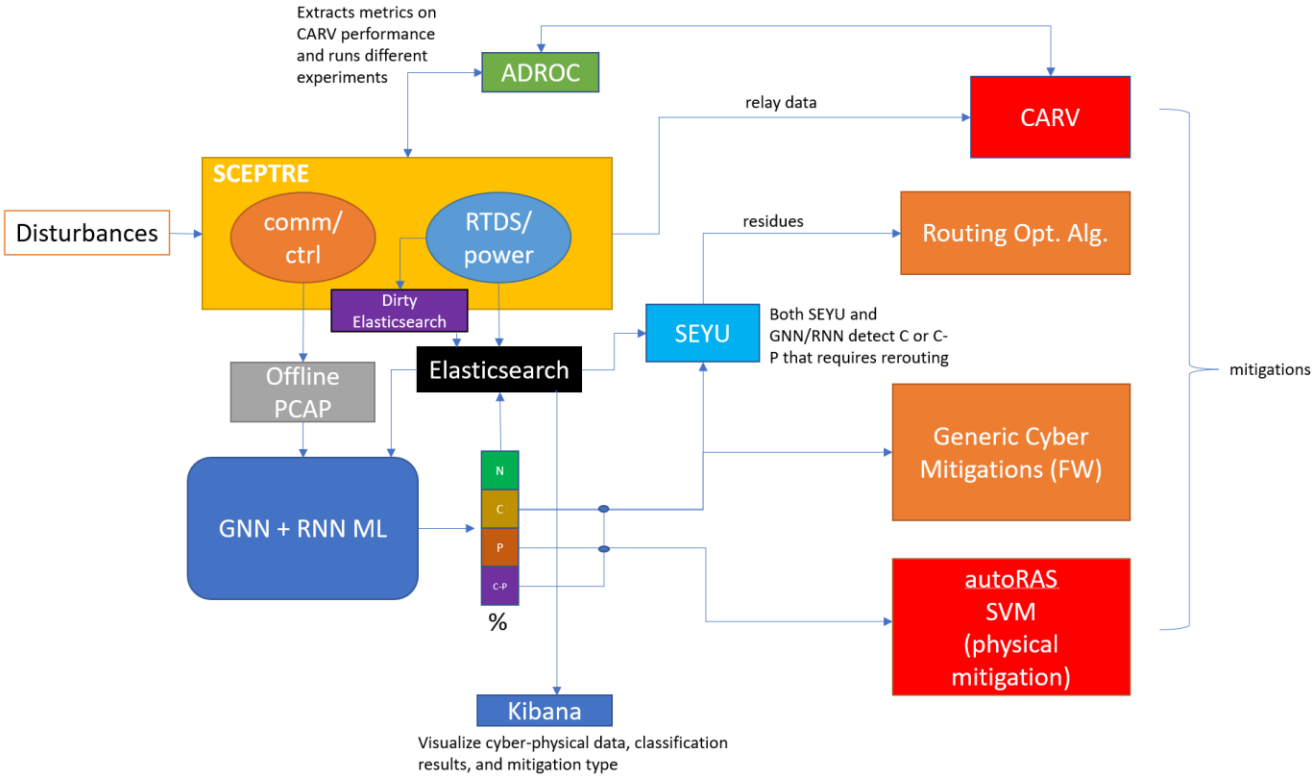


Figure 34: Emulation environment components to support FDI experiment using IEEE 39-bus system use-case.

For this FDI scenario, the IEEE 39-bus system and associated synthetic communication network were implemented within the emulation. The one-line diagram for the IEEE 39-bus system is shown in Figure 35 and the associated SCEPTRE™ topology is shown in Figure 36.

To sample values, 8 PMUs were placed in the RSCAD model of the IEEE 39-bus system, using an optimal PMU placement method described in [65], sampling a subset of the buses. The reason only a subset was sampled was due to a hardware limitation with the RTDS GTNET card that restricted the total number of PMUs to 8. Then, 8 RTUs in the emulation environment were configured to receive data from the 8 PMUs configured in the RTDS via C37.118 protocol. The data from the RTUs is sent to the SCADA server via the DNP3 protocol. The SCADA server is an open platform communications (OPC) server running Software Toolbox TOP Server version 5 [66].

To collect data that was modified with the FDI, an additional Elasticsearch server was added for the 39-bus topology. This "dirty" Elasticsearch server resided within the emulation environment to collect "dirty" data as the SCADA server observed it, including any modifications that were made by the FDI scenario (shown in Figure 38). The "dirty" data was collected by a Python script that queried the OPC server via the OPC-UA protocol, using the asyncua library [67], and sent the collected data to the "dirty" Elasticsearch instance using the elasticsearch-py library [68].

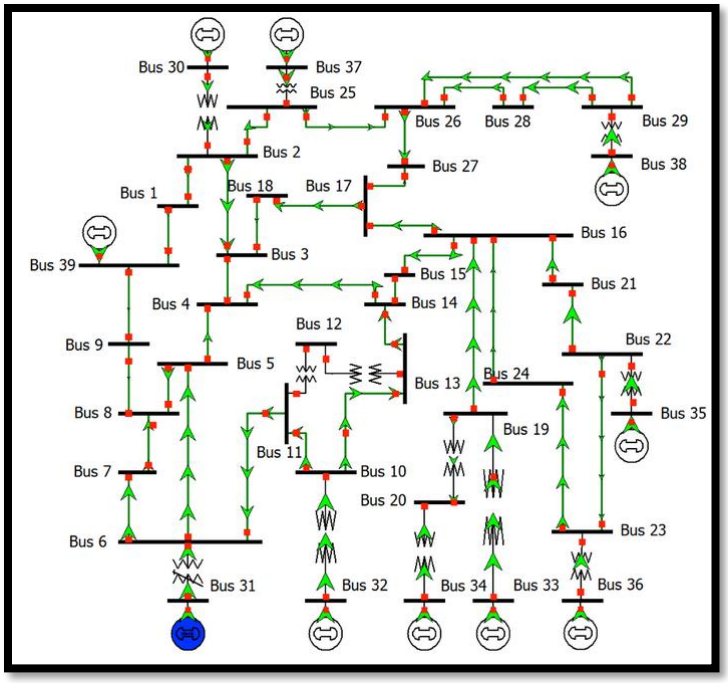


Figure 35: IEEE 39-bus system one-line diagram.

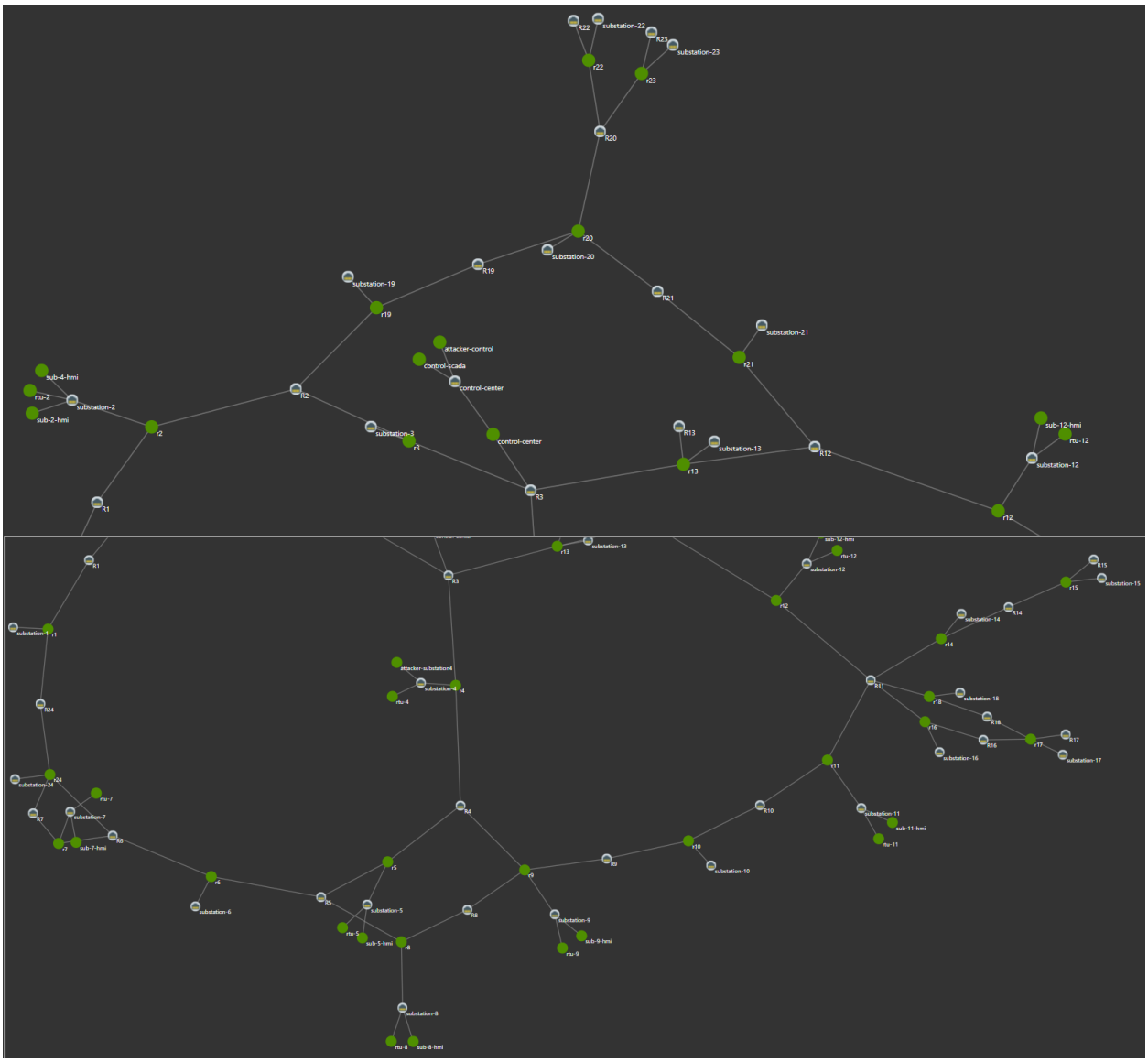


Figure 36: SCEPTRE™ topology for IEEE 39-bus system.

7.2.1. Scenario Description

The stealthy FDI attack developed cannot be detected by traditional BDD tests, and thus requires evaluation of the physical data streams and the fastSE implementation to detect it. For this experiment, the stealthy FDI is categorized as a cyber-physical disturbance due to the data injection activity (although not detected by traditional BDD tests) and impact to the physical data. In this experiment, the team tested the detection of the FDI using fastSE, the real-time state estimator developed by the TAMU team; the FDI is targeting substation 4 in Figure 38. In this scenario, no major power system consequences resulted, only small variations were introduced to the power system data from the FDI. Therefore, this scenario is focused on reconnaissance activity from an adversary testing their FDI capabilities and gathering further information about the system.

As discussed in the previous section, this stealthy FDI scenario was implemented in the RTDS and SCEPTRE™ emulation environment using the IEEE 39-bus system use-case. Updates to the emulation environment included implementation of fastSE, a corrupted, 'dirty' Elasticsearch implementation (with data corrupted with the FDI), and routing optimization algorithm.

7.2.2. Mitigation

In this experiment, if the FDI is detected by fastSE, the routing optimization algorithm described in Section 5.4 is launched. The adaptive routing algorithm is developed to actively optimize the routing path of the communication system, based on the feedback from the state estimator. Specifically, the residues from the state estimator are used to detect the FDI and inform the new routing architecture. For the compromise of substation 4 with the FDI, its high residue will indicate the FDI attack and cause the routes to be updated to avoid it.

7.2.3. Results and Next Steps

The FDI attack was implemented within the IEEE 39-bus system use-case emulation, targeting substation 4. The necessary scenario components were also incorporated into the emulation environment, including: FDI attack script, fastSE, dirty Elasticsearch, and routing optimization. However, we ran into a few challenges with adapting fastSE to the RTDS environment and its function in detecting the FDI.

fastSE, like other transmission system state estimators, is developed for single-phase positive sequence steady state power system analysis. When used with RTDS, certain challenges raise, and they can be categorized into mainly four aspects:

Simulation type

- RTDS is known for its capability to conduct electromechanical transient simulation (TS) and electromagnetic transient simulation (EMT). TS and EMT are normally formulated as a set of partial differential equations (PDEs), and they tend to have higher convergence tolerance and thus are more sensitive to initial states, compared to power flow which is usually formulated as a set of ordinary differential equations. Two major causes for the mismatch between RTDS and PowerWorld (PW) results in this category can be the initial guess on system states, and the initial voltage and frequency. PDEs may converge to a different equilibrium point even using the same initial guess as used in the power flow.

Model

- System models can be different in RTDS compared to PW. This has been well discussed in this paper [69]. One thing to specifically point out is that if the load model is constant impedance, setting the maximum and minimum power to be the same as nominal power may introduce unexpected errors, cause the constant impedance load has varying power, while conversely the constant power load has varying impedance.

Case

- RTDS cannot output either the admittance matrix or the impedance matrix, so it is very hard to validate whether the case is the same as in the PW. Besides, in the standard model, the power base and the voltage base are different compared to PW.

Data

- Considering the data is coming from the simulated PMU channels, there is also possibility that either the probing location or the PMU settings are mistaken. Moreover, it is also important to make sure that the correct formula is used to convert the three phase measurements to single phase positive sequence values.

With additional time, the team may have been able to overcome these aforementioned challenges to enable the successful function of fastSE in our emulation environment. We are planning on working through these issues in future work. For this particular experiment and our time constraints, we decided to whitecard the use of fastSE to detect the FDI attack. Thus, we focused on testing the adaptive routing optimization as a mitigation against the FDI attack within the emulation environment. As can be seen in Figure 37, the FDI attack corrupted the SCADA data output by the RTDS for the IEEE 39-bus system.

Figure 37: Screenshot of Kibana dashboards highlighting the FDI attack script success by corrupting the SCADA data from the RTDS (right figure) from the original, "clean" data (left figure).

The adaptive routing optimization algorithm was then launched and was successfully able to isolate the compromised substation 4 and reroute traffic, as shown in Figure 38. In this manner, communications and data flows can continue without corruption from the FDI attack and the compromised substation can be isolated and mitigated with upstream intervention (e.g., after notification of utility and/or other cybersecurity defenses).

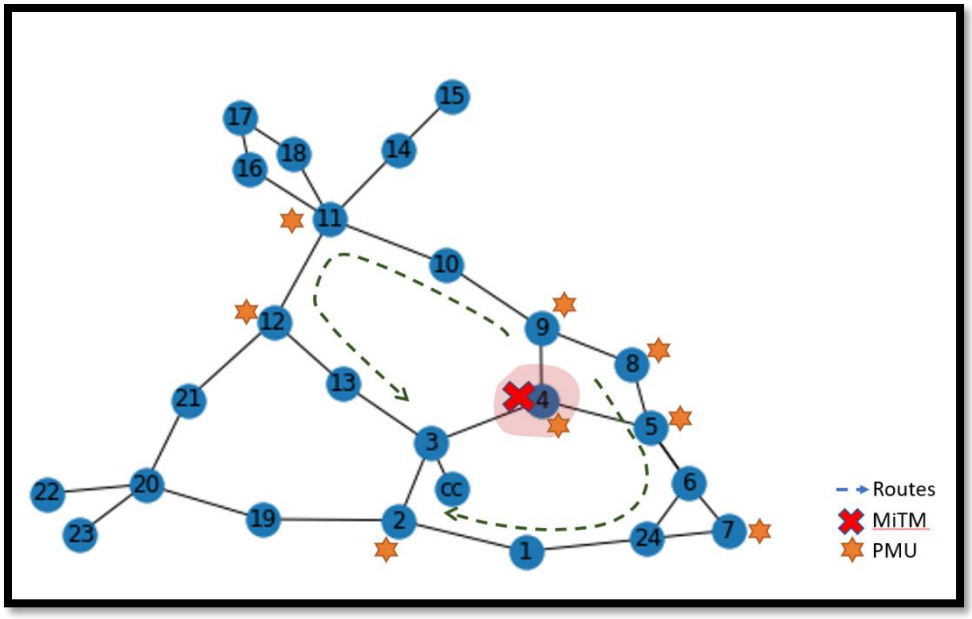


Figure 38: Result of adaptive routing optimization algorithm in isolating Substation 4 and rerouting communication traffic.

Lastly, due to time constraints and focus on testing the routing optimization mitigation, we did not test the HARMONIE-SPS ML methodology against the collect FDI data. However, we provide some discussion on the differences/challenges to address when moving to a larger scale grid system.

Moving from the simplified 9-bus system to a larger and more complicated 39-bus system (with its larger and more complex associated cyber infrastructure) has some potential implications for how we consider 'scaling up' the generation of ML datasets and adapting the architecture and training of the ML model. The primary challenges associated with adapting machine-learning approaches to more 'complex' datasets are:

1. Data dimensionality – The input data for ML is more complex and generally of higher 'dimension'. This creates 2 potential issues. First, the ML model will require additional trainable parameters to encode the higher-dimension data. Increasing the 'complexity' of the ML model (which is roughly correlated with trainable-parameter count) typically leads to a requirement for additional training data samples in order to avoid 'overfitting' the more-complex model to limited training data. Second, the well-studied 'curse of dimensionality' suggests that the requirement for additional data samples from higher-dimensional spaces is fundamental and not only tied to model overfitting problems. In general, we would expect the number of data samples required to adequately capture the variance associated with higher-dimensional data to increase approximately-exponentially with increased data dimensionality. So, increasing the 'complexity' of the dataset even a little bit is expected to increase the number of data samples required to train an ML model by a lot.
2. Sampling bias – In an ideal world, ML models would be trained using unbiased or 'random' sampling from the appropriate 'population' distribution of data relevant to the problem of interest. Any potential 'bias' in data sampling may negatively impact the trained model's ability to 'generalize' or perform well in practice. When training data are simulated or emulated, it may be challenging to engineer procedures that are: 1) scalable to very large

sample sizes (millions), which typically requires a high degree of automation, 2) random with respect to the population of data on which the model is expected to perform, which may be unknown, and 3) realistic in terms of generating training data that represent real-world scenarios. In many cases, generating 'realistic' training data necessary to train an ML model to perform well 'in the real world' is a time-consuming process that doesn't scale to the generation of millions of training samples or – if it does scale well – fails to generalize to capture the 'entire' population of relevant data. In our view, the development of scalable, realistic data generation procedures is a high-priority if we want to be able to train reliable ML models under complex scenarios.

State-of-the-art ML models like deep neural networks trained using supervised-learning have been shown to perform exceptionally well in complex inference tasks when:

1. There is 'enough' training data
2. The inference task is well-defined at training time

In practical terms, a 'well-defined' inference task for a machine-learning classifier means that the set of 'ground-truth' class labels and input data examples associated with each class label are bounded and known a priori. The ML model is trained to extract and associate relevant 'features' from the input data with the known class-label of each data sample. The procedure assumes that the collected set of input data is sufficient to identify all the relevant features that will be associated with each class label at training time, and that all of the relevant class labels are also known at training time. Model performance is difficult to predict in cases in which 'novel' features associated with an existing class label arise after training. Similarly, 'adapting' a trained model to accurately predict novel class labels while retaining good performance on its existing classification tasks is not always trivial.

In contrast to this paradigmatic expectation, we generally expect 'disturbance-event' detection and mitigation paradigms to be an 'evolving' or 'dynamic' landscape of data and associated class labels. The classic example is the 'zero-day' event, in which a completely 'novel feature' associated with a disturbance event arises. Under this scenario, the performance of a previously-trained ML model would be difficult to predict. Similarly, a complex scenario in which various (partially effective) mitigation strategies are actively deployed during the course of e disturbance event may strongly impact the performance of an ML model trained without seeing mitigation strategy deployment. This type of situation creates both practical and fundamental issues. More fundamentally, what is the 'desired ground-truth' output of a trained ML algorithm under a scenario in which the state of the managed system is neither 'okay' nor 'not-okay'?

Our current work considered only supervised-learning ML paradigms, but it may be that alternative paradigms (e.g., reinforcement-learning, open-set classification) may be more appropriately suited to the ultimate goals of the project. We suggest that future work would benefit from an investigation into relevant deployment scenarios to define specific requirements and an evaluation of ML paradigms to determine which paradigms are most likely to meet those requirements.

7.3. Evaluation of CARV Scheme

To demonstrate the application of this voting scheme and examine how the CARV algorithm originally developed in [50] can be tailored for a RAS, we will integrate it with a load shedding scheme developed for the WSCC 9-bus system, which is shown in Figure 24.

The protection scheme to be integrated with CARV consists of shedding load to alleviate line overloading and under-voltage issues that arise from a contingency that consists of the loss of

generation at bus 1 and a disconnect on the branch between buses 6 and 9. This protection scheme was developed from the methodology presented in [46], and is aimed to mitigate the overloading conditions that develop in the system due to this contingency by shedding load in the system.

Unsurprisingly, the main effect of this contingency is overloading of the line between buses 5 and 7 (line 5-7), where large amounts of current flow due to the fault in line 6-9, as well as voltage drops in the system, most significantly at bus 6. To reduce these issues and allow the system at large to stay operational, the active and reactive power requirements at buses 5 and 6 are reduced significantly by the RAS by shedding load.

In order to test the resilience benefits of CARV, several scenarios will be tested. All cases here will be used to test the appropriate arming and triggering of the applied SPS, and for comparison we will apply all cases to the SPS without any relay voting and also with the relay voting scheme applied to both investigate the additional overhead that comes from the added communications but also where there may be benefits or drawbacks.

The scenarios that were implemented are:

1. Base case, contingency with no relay issues
2. Branch 8-9 relay unresponsive, or lies about conditions and actions
3. Load 6 unresponsive, or lies about its conditions and actions
4. Relay x & y unresponsive (multiple, different groups)
5. Relay x & z unresponsive (multiple, same group)
6. 3 relays unresponsive, same voting group

For the load shedding scheme that distributed consensus algorithm is implementing, the calculation is calculating the total loads in the two control areas used for the SPS. Note that this requires *a-priori* knowledge of the number of sensors recording loads in the system in order to convert from an average value to the total value, which in this instance are the 3 relays connected to the loads in the system.

There are various configuration parameters that impact CARV performance, which are detailed further in the full paper [51]. Figure 39 provides an example for one of the configuration parameters, p, which corresponds to the system putting more weight on the existing consensus value than new measurements.

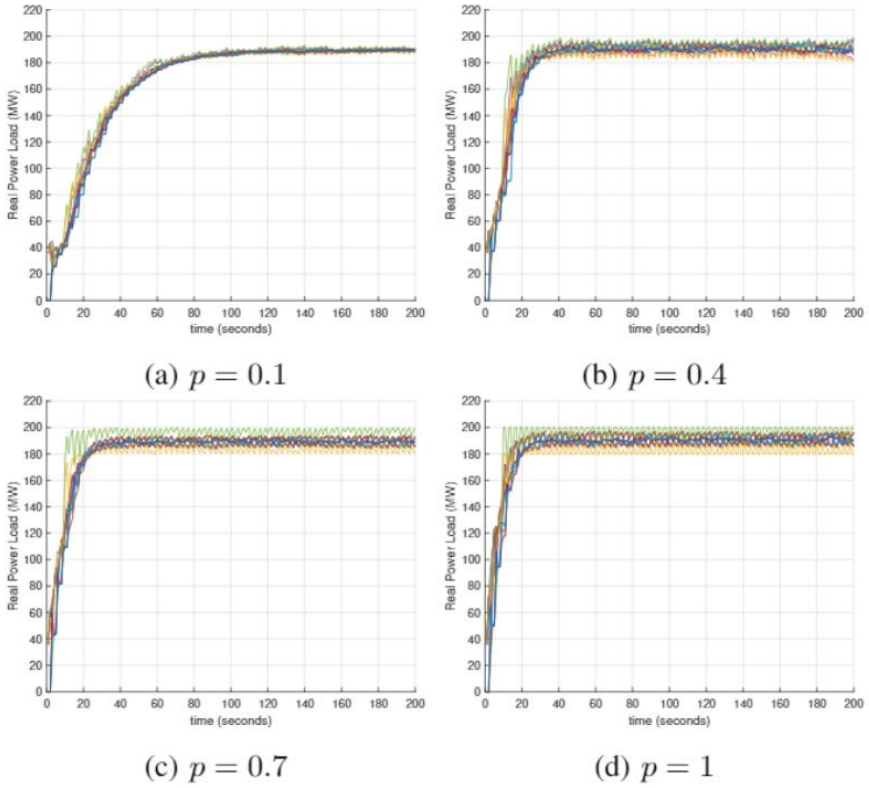
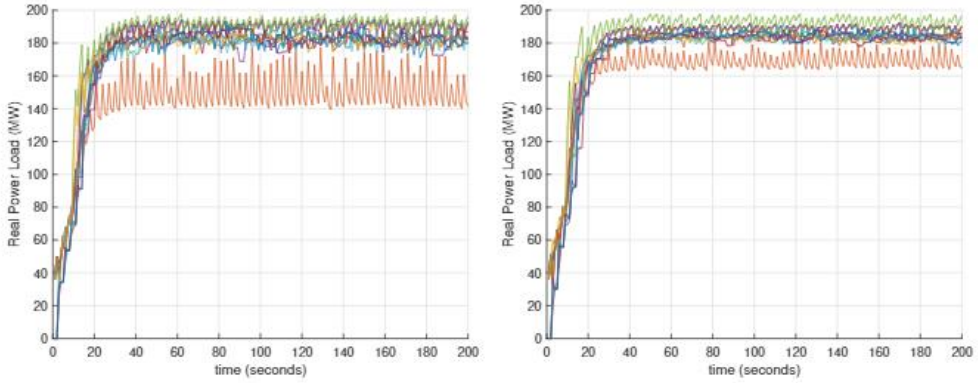


Figure 39: Consensus load in control area 1 for various averaging factors p.

In Figure 40, we can see what occurs when a relay is misreporting values but is detected and flagged by the other relays in the system. This is done by having each relay check if the reported values by the other relays are within a certain threshold of its own current value before doing a new calculation. In this instance, that flagging threshold is set to ±10% of the relays current value. Note that exceptions need to be included for sensors in the system that are adding new measurements, and this procedure needs to be incorporated both into the regular voting group consensus calculations and the inter-group communications as well, so there are limitations to this type of detection mechanism. However, as we can see in this result, each relay is able to check and see that the branch 8-9 relay is reporting its value and isolate it. Additionally, this can serve as a detection mechanism if a relay stops communicating with its peers. In this procedure, if a relay is flagged then the other relays in the system will independently temporarily remove that misbehaving relay from the consensus algorithm process until it could be checked and fixed and can report the issue to a control center or other central control system.



(a) Branch 8-9 relay reports load as 60% of estimated value

(b) Branch 8-9 relay reports load as 80% of estimated value

Figure 40: Consensus load when relay at Branch 8-9 misreports values.

Additional results from the different scenarios explored to evaluate CARV are found in the full paper [51].

8. CONCLUSIONS AND FUTURE WORK

The HARMONIE-SPS methodology developed under this LDRD project is able to:

- Classifies system conditions using transformer-based and graph convolution neural networks
- Deploys cyber-physical mitigations depending on classification
    - Leverages autoRAS for automating corrective action and triggering condition pairs, CARV for improved relay operation, and routing optimization for continued communications

All in all, HARMONIE-SPS was able to demonstrate how processing both cyber-physical data for assessing system conditions and informing response provides more comprehensive, effective protection of the electric grid. This is especially relevant as more and more smart devices are added to the grid and distributed systems such as DERs increase their penetration. Furthermore, the use of a cyber-physical emulation environment with the integration of RTDS and SCEPTRE™ exemplified the need for real-time environments for training and testing machine learning based mechanisms and for exploring adaptive use-cases. Adaptive applications/tools such as autoRAS, CARV, and the routing optimization algorithm utilized the real-time, cyber-physical data streams and showed how real-time response and built-in fault tolerance (i.e., within CARV) is necessary for more resilient operation of the grid during disturbances.

For future development of the HARMONIE-SPS methodology, we would like to iterate on more sophisticated disturbances within the emulation environment such as the FDI attack. At the end of this project, we were able to successfully implement and utilize the FDI script, dirty Elasticsearch instance, and adaptive routing optimization algorithm; we faced challenges adapting fastSE to the RTDS and SCEPTRE™ environment and would like to resolve this in the future. We additionally would like to test disturbances such as extreme weather events coupled with cyber attacks (e.g., multi-hazard events) and test with larger power system use-cases.

More complex communication networks would also be of interest to analyze the properties of collected cyber-physical data. With increased amount and complexity of this data, it is important to test the HARMONIE-SPS ML framework robustness under different data availability scenarios to understand the limits of its classification. Additionally, we can explore the use of further OOB data sources such as IDS alerts, including the team's related proactive intrusion detection and mitigation system (PIDMS) research [70]. Further future work in exploring alternative ML paradigms are described in Section 7.2. Testing with field data and in partnership with utilities would also inform realistic data availability and formats for HARMONIE-SPS ingestion. This utility partnership would also start exploration of different HARMONIE-SPS deployment options (e.g., standalone tool augmenting existing SPSs or adaptive SPS integrated with playbooks).

Lastly, expanding the suite of cyber-physical corrective actions to include more sophisticated control strategies and/or cybersecurity defenses is of interest. For example, utilizing DERs to respond to generation loss and/or instabilities would expand HARMONIE-SPS to account for system architecture, DER penetration levels, and instability metrics. On the cyber-side, techniques such as moving target defense to thwart adversaries could be added and how this may or may not affect SPS function could be explored. Another aspect that was not quite explored in this project iteration is the prioritization of speed, selectivity and security using the ML classification results and informing cyber-physical response. With a larger suite of response actions and disturbance scenarios, this prioritization process is of great interest for future HARMONIE-SPS development.

REFERENCES

[1] K. Hemsley and R. Fisher, "History of Industrial Control System Cyber Incidents," Idaho National Laboratory, 2018.

[2] "Framework for Improving Critical Infrastructure Cybersecurity," National Institute of Standards and Technology, 2018.

[3] P. Mazzei, I. Penn, F. Robles, "With Earthquakes and Storms, Puerto Rico's Power Grid Can't Catch a Break," *N. Y. Times*, 2020.

[4] "Analysis of the cyber attack on the Ukrainian power grid," *Electr. Inf. Shar. Anal. Cent. E-ISAC*, vol. 388, 2016.

[5] M. Törngren and P. T. Grogan, "How to Deal with the Complexity of Future Cyber-Physical Systems?," *Designs*, vol. 2, no. 4, Art. no. 4, Dec. 2018, doi: 10.3390/designs2040040.

[6] C. Lai, N. Jacobs, S. Hossain-McKenzie, P. Cordeiro, O. Onunkwo, and J. Johnson, "Cyber Security Primer for DER Vendors, Aggregators, and Grid Operators," Sandia National Laboratories, Sandia Report SAND2017-13113, Dec. 2017.

[7] "Remedial Action Scheme Design Guide." Western Electricity Coordinating Council (WECC), 2011.

[8] P. Anderson and B. LeReverend, "Industry experience with special protection schemes," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1166–1179, 1996.

[9] J. G. O'Brien, E. L. Barrett, X. Fan, R. Diao, R. Huang, and Q. Huang, "Survey on RAS/SPS modeling practice," Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2017.

[10] M. D. Maram and N. Amjady, "Event-based remedial action scheme against super-component contingencies to avert frequency and voltage instabilities," *IET Gener. Transm. Distrib.*, vol. 8, no. 9, pp. 1591–1603, 2014.

[11] S. Wang and G. Rodriguez, "Smart RAS (Remedial Action Scheme)," in *2010 Innovative Smart Grid Technologies (ISGT)*, 2010, pp. 1–6.

[12] Y. Zhang and K. Tomsovic, "Adaptive remedial action scheme based on transient energy analysis," in *IEEE PES Power Systems Conference and Exposition, 2004.*, 2004, pp. 925–931.

[13] H. Atighechi, P. Hu, J. Lu, G. Wang, and S. Ebrahimi, "A fast load shedding remedial action scheme using real-time data for BC hydro system," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[14] X. Fan *et al.*, "Adaptive RAS/SPS System Setting for Improving Grid Reliability and Asset Utilization through Predictive Simulation and Controls: A Use Case for Transformative Remedial Action Scheme Tool (TRAST): Jim Bridger RAS Evaluation and Analysis," Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2019.

[15] S. Hossain-McKenzie, K. Davis, M. Kazerooni, S. Etigowni, S. Zonouz, "Distributed controller role and interaction discovery," 2017.

[16] S. Hossain-McKenzie, "Protecting the power grid: strategies against distributed controller compromise," PhD Thesis, University of Illinois Urbana-Champaign, 2017.

[17] S. Boyd, N. Parikh, C. Chu, B. Peleato, J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, 2011.

[18] T. R. Camacho-Lopez, "SCEPTRE.," Aug. 2016, [Online]. Available: https://www.osti.gov/biblio/1376989

[19] R. Kuffel, J. Giesbrecht, T. Maguire, R. P. Wierckx, and P. McLaren, "RTDS-a fully digital power system simulator operating in real time," in *Proceedings 1995 International Conference on*

Energy Management and Power Delivery EMPD '95, 1995, vol. 2, pp. 498–503 vol.2. doi: 10.1109/EMPD.1995.500778.
[20] The MITRE Corporation, "ATT&CK," 2020, [Online]. Available: https://attack.mitre.org/
[21] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, 2017, vol. 30. Accessed: Jul. 28, 2022. [Online]. Available: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.
[23] T. B. Brown *et al.*, "Language Models are Few-Shot Learners." arXiv, Jul. 22, 2020. doi: 10.48550/arXiv.2005.14165.
[24] B. Tang and D. S. Matteson, "Probabilistic Transformer For Time Series Analysis," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 23592–23608. Accessed: Jul. 31, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/c68bd9055776bf38d8fc43c0ed283678-Abstract.html
[25] R. Xiong *et al.*, "On Layer Normalization in the Transformer Architecture." arXiv, Jun. 29, 2020. doi: 10.48550/arXiv.2002.04745.
[26] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The Long-Document Transformer." arXiv, Dec. 02, 2020. doi: 10.48550/arXiv.2004.05150.
[27] N. Kitaev, Ł. Kaiser, and A. Levskaya, "Reformer: The Efficient Transformer." arXiv, Feb. 18, 2020. doi: 10.48550/arXiv.2001.04451.
[28] Y. Xiong *et al.*, "Nystr\"omformer: A Nystr\"om-Based Algorithm for Approximating Self-Attention." arXiv, Mar. 31, 2021. doi: 10.48550/arXiv.2102.03902.
[29] M. Zaheer *et al.*, "Big Bird: Transformers for Longer Sequences." arXiv, Jan. 08, 2021. doi: 10.48550/arXiv.2007.14062.
[30] D. Calzada, S. Hossain-McKenzie, and Z. Mao, "Deep Learning Architecture for Processing Cyber-Physical Data in the Electric Grid," in *2022 IEEE Power and Energy Conference at Illinois (PECI)*, Mar. 2022, pp. 1–8. doi: 10.1109/PECI54197.2022.9744015.
[31] A. Nicolae, "PLU: The Piecewise Linear Unit Activation Function." arXiv, Sep. 03, 2018. doi: 10.48550/arXiv.1809.09534.
[32] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization." arXiv, Jan. 04, 2019. doi: 10.48550/arXiv.1711.05101.
[33] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay." arXiv, Apr. 24, 2018. doi: 10.48550/arXiv.1803.09820.
[34] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks." arXiv, Apr. 04, 2017. doi: 10.48550/arXiv.1506.01186.
[35] E. D. Cubuk, B. Zoph, J. Shlens, and Q. Le, "RandAugment: Practical Automated Data Augmentation with a Reduced Search Space," in *Advances in Neural Information Processing Systems*, 2020, vol. 33, pp. 18613–18624. Accessed: Jul. 31, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/d85b63ef0ccb114d0a3bb7b7d808028f-Abstract.html
[36] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, "Recent Advances in Adversarial Training for Adversarial Robustness." arXiv, Apr. 20, 2021. doi: 10.48550/arXiv.2102.01356.
[37] R. Müller, S. Kornblith, and G. Hinton, "When Does Label Smoothing Help?" arXiv, Jun. 10, 2020. doi: 10.48550/arXiv.1906.02629.

[38] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards Deep Learning Models Resistant to Adversarial Attacks," presented at the International Conference on Learning Representations, Mar. 2022. Accessed: Jul. 31, 2022. [Online]. Available: https://openreview.net/forum?id=rJzIBfZAb
[39] E. Wong, L. Rice, and J. Z. Kolter, "Fast is better than free: Revisiting adversarial training." arXiv, Jan. 12, 2020. doi: 10.48550/arXiv.2001.03994.
[40] G. Sriramanan, S. Addepalli, A. Baburaj, and V. B. R, "Towards Efficient and Effective Adversarial Training," in *Advances in Neural Information Processing Systems*, 2021, vol. 34, pp. 11821–11833. Accessed: Jul. 31, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/62889e73828c756c961c5a6d6c01a463-Abstract.html
[41] T. N. Kipf and M. Welling, "Semi-Supervised Classification with Graph Convolutional Networks." arXiv, Feb. 22, 2017. doi: 10.48550/arXiv.1609.02907.
[42] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks." arXiv, Mar. 29, 2019. doi: 10.48550/arXiv.1812.04948.
[43] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN." arXiv, Mar. 23, 2020. doi: 10.48550/arXiv.1912.04958.
[44] F. Chollet, "Xception: Deep Learning With Depthwise Separable Convolutions," 2017, pp. 1251–1258. Accessed: Aug. 03, 2022. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2017/html/Chollet_Xception_Deep_Learning_CVPR_2017_paper.html
[45] D. Haase and M. Amthor, "Rethinking Depthwise Separable Convolutions: How Intra-Kernel Correlations Lead to Improved MobileNets," Jun. 2020, pp. 14588–14597. doi: 10.1109/CVPR42600.2020.01461.
[46] H. Li, K. Shetye, T. Overbye, K. Davis, and S. Hossain-Mckenzie, *Towards the Automation of Remedial Action Schemes Design*. 2021. Accessed: Sep. 02, 2022. [Online]. Available: http://hdl.handle.net/10125/71025
[47] B. Thayer, Z. Mao, Y. Liu, K. Davis, and T. Overbye, "Easy SimAuto (ESA): A Python Package that Simplifies Interacting with PowerWorld Simulator," *J. Open Source Softw.*, vol. 5, p. 2289, Jun. 2020, doi: 10.21105/joss.02289.
[48] H. Li, K. S. Shetye, S. Hossain-McKenzie, K. Davis, and T. J. Overbye, "Investigation of Automated Corrective Actions for Special Protection Schemes," Sandia National Lab. (SNL-NM), Albuquerque, NM (United States); Texas A & M Univ., College Station, TX (United States), SAND2020-9602, Sep. 2020. doi: 10.2172/1668137.
[49] H. J. Altuve, K. Zimmerman, and D. Tziouvaras, "Maximizing line protection reliability, speed, and sensitivity," in *2016 69th Annual Conference for Protective Relay Engineers (CPRE)*, 2016, pp. 1–28. doi: 10.1109/CPRE.2016.7914896.
[50] N. Jacobs *et al.*, "Next-Generation Relay Voting Scheme Design Leveraging Consensus Algorithms," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*, Apr. 2021, pp. 1–6. doi: 10.1109/PECI51586.2021.9435201.
[51] N. Jacobs, S. Hossain-McKenzie, and A. Summers, "Consensus Algorithm Relay Voting (CARV) for Improving Resilience in Grid Cybersecurity," *IEEE Trans. Smart Grid*, vol. [To Be Submitted], Sep. 2022.
[52] M. Castro and B. Loskov, "Practical Byzantine Fault Tolerance," New Orleans, USA, Feb. 1999.
[53] P. Aublin, S. B. Mokhtar, and V. Quéma, "RBFT: Redundant Byzantine Fault Tolerance," in *2013 IEEE 33rd International Conference on Distributed Computing Systems*, 2013, pp. 297–306. doi: 10.1109/ICDCS.2013.53.

[54] A. Sahu *et al.*, "Design and evaluation of a cyber-physical testbed for improving attack resilience of power systems," *IET Cyber-Phys. Syst. Theory Appl.*, vol. n/a, no. n/a, doi: https://doi.org/10.1049/cps2.12018.
[55] P. Wlazlo *et al.*, "Man-in-the-middle attacks and defence in a power system cyber-physical testbed," *IET Cyber-Phys. Syst. Theory Appl.*, vol. 6, no. 3, pp. 164–177, 2021, doi: 10.1049/cps2.12014.
[56] M. Soetan, Z. Mao, and K. Davis, "Statistics for Building Synthetic Power System Cyber Models," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*, Apr. 2021, pp. 1–5. doi: 10.1109/PECI51586.2021.9435196.
[57] A. Summers, C. Goes, D. Calzada, N. Jacobs, S. Hossain-McKenzie, and Z. Mao, "Towards Cyber-Physical Special Protection Schemes: Design and Development of a Co-Simulation Testbed Leveraging SCEPTRE™," in *2022 IEEE Power and Energy Conference at Illinois (PECI)*, Mar. 2022, pp. 1–7. doi: 10.1109/PECI54197.2022.9744043.
[58] "CALDERA." https://caldera.mitre.org/ (accessed Sep. 07, 2022).
[59] "Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic." https://www.elastic.co/ (accessed Sep. 07, 2022).
[60] S. Hossain-McKenzie *et al.*, "Adaptive, Cyber-Physical Special Protection Schemes to Defend the Electric Grid Against Predictable and Unpredictable Disturbances," in *2021 Resilience Week (RWS)*, 2021, pp. 1–9.
[61] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[62] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *ArXiv Prepr. ArXiv14123555*, 2014.
[63] N. Geneva and N. Zabaras, "Transformers for modeling physical systems," *ArXiv Prepr. ArXiv201003957*, 2020.
[64] Z. Mao, "🚀FastSE." Aug. 03, 2022. Accessed: Sep. 07, 2022. [Online]. Available: https://github.com/mzy2240/fastSE
[65] V. V. R. Raju and S. V. J. Kumar, "An optimal PMU placement method for power system observability," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, Feb. 2016, pp. 1–5. doi: 10.1109/PECI.2016.7459248.
[66] S. Toolbox, "TOP Server - Industrial OPC Connectivity for 100's of Devices." https://products.softwaretoolbox.com/top-server/opc-da-ua-suitelink-drivers (accessed Sep. 20, 2022).
[67] "opcua-asyncio." Free OPC-UA Library, Sep. 19, 2022. Accessed: Sep. 20, 2022. [Online]. Available: https://github.com/FreeOpcUa/opcua-asyncio
[68] "Elasticsearch Python Client." elastic, Sep. 20, 2022. Accessed: Sep. 20, 2022. [Online]. Available: https://github.com/elastic/elasticsearch-py
[69] W. Trinh, Z. Mao, T. Overbye, J. Weber, and D. Morrow, "Considerations in the Initialization of Power Flow Solutions from Dynamic Simulation Snapshots," Apr. 2021, pp. 1–6. doi: 10.1109/NAPS50074.2021.9449750.
[70] S. Hossain-McKenzie, A. Chavez, N. Jacobs, C. Jones, A. Summers, and B. Wright, "Securing Inverter Communication: Proactive Intrusion Detection and Mitigation System to Tap, Analyze, and Act.," Sandia National Lab. (SNL-NM), Albuquerque, NM (United States), SAND2022-3759, Mar. 2022. doi: 10.2172/1861984.

APPENDIX A. SUPPLEMENTAL MACHINE LEARNING RESULTS

Figure A-1, Figure A-2, and Figure A-3-- training run traces (Figure 4) for replicates 2-4:

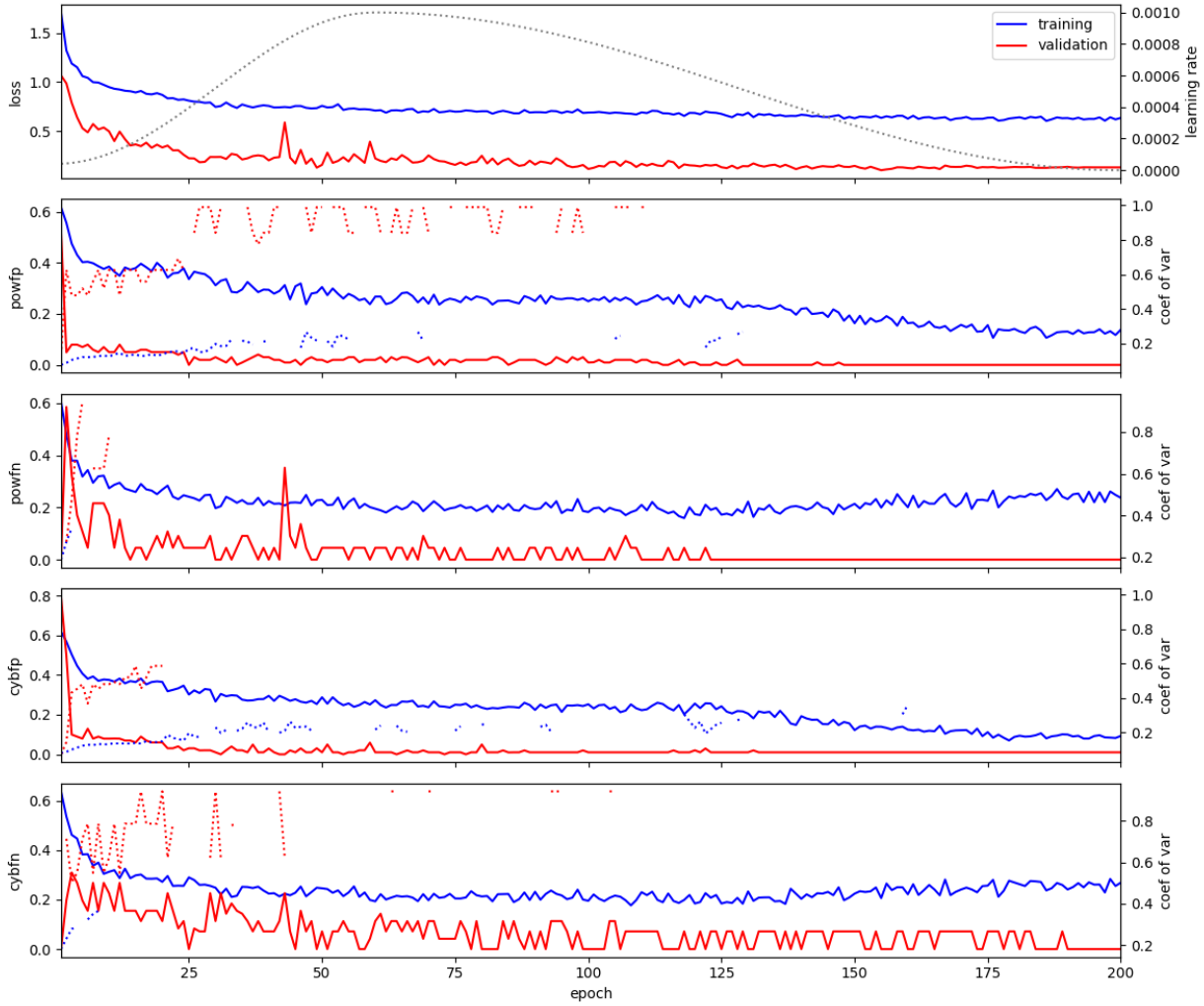


Figure A- 1

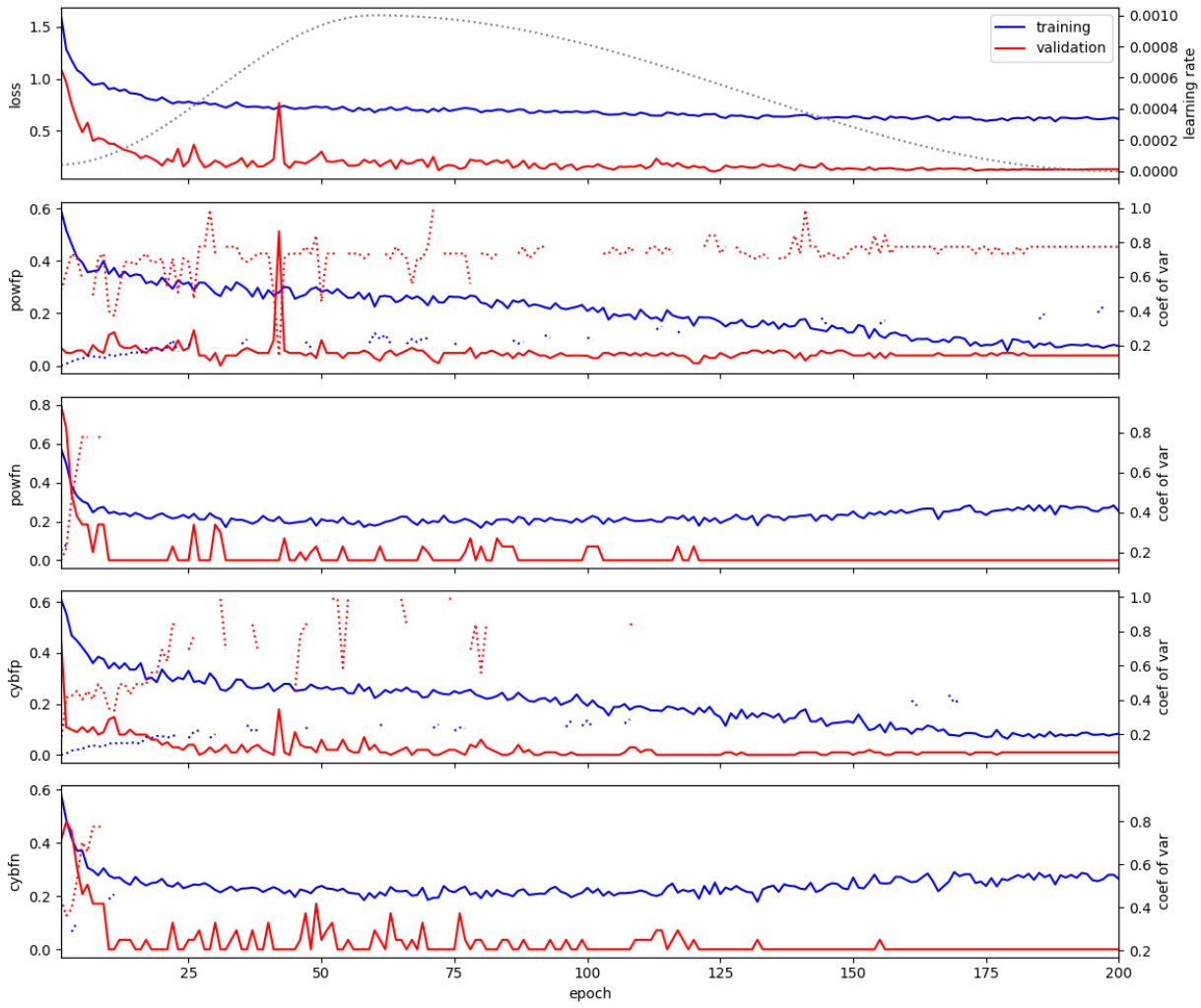

training
validation
loss
learning rate
powfp
coef of var
powfn
coef of var
cybfp
coef of var
cybfn
coef of var
epoch

Figure A- 2

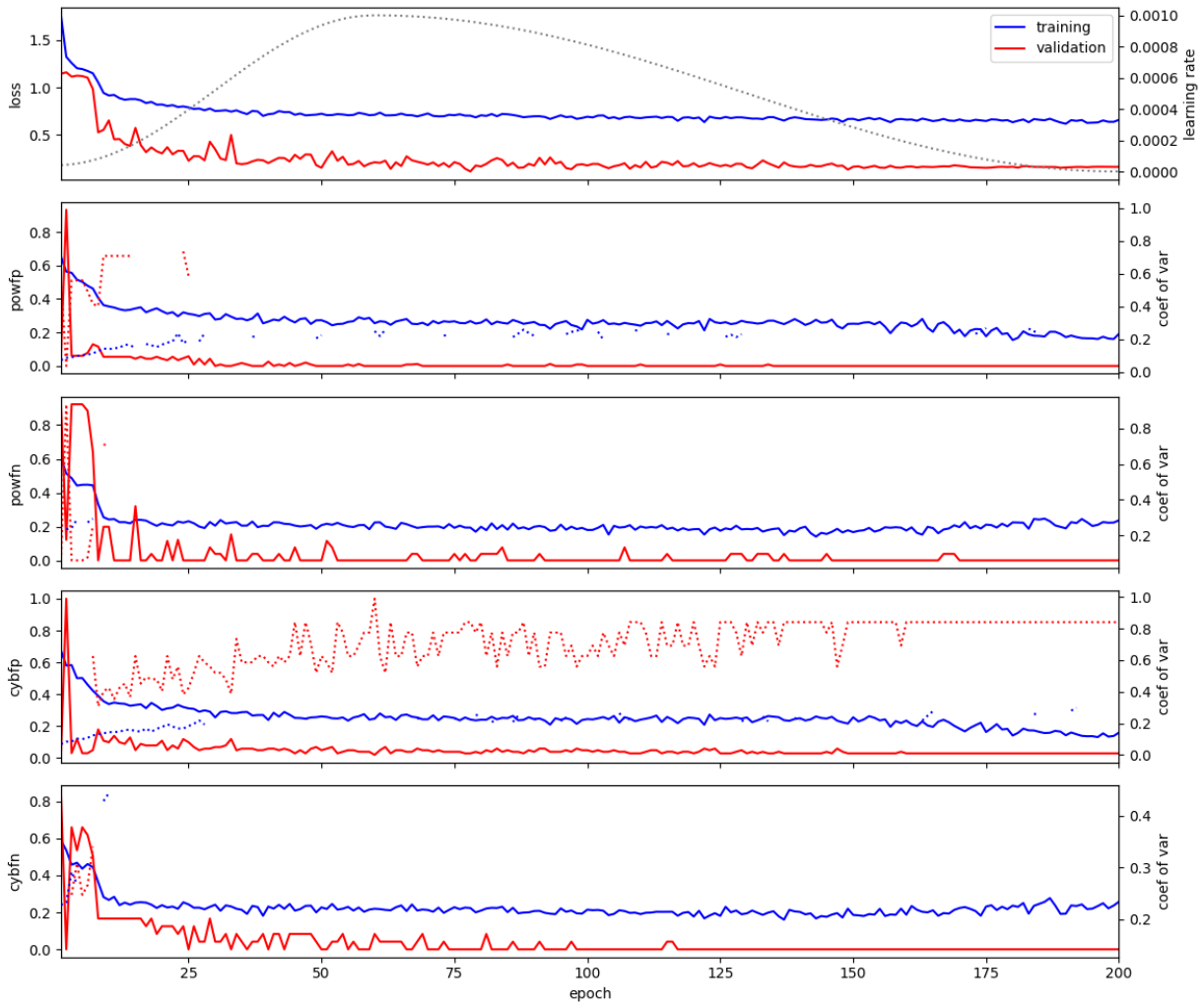

training
validation
loss
learning rate
powfp
powfn
cybfp
cybfn
coef of var
epoch

Figure A- 3

Figure A-4, Figure A-5, and Figure A-6-- ROC curves (Figure 5) for replicates 2-4:

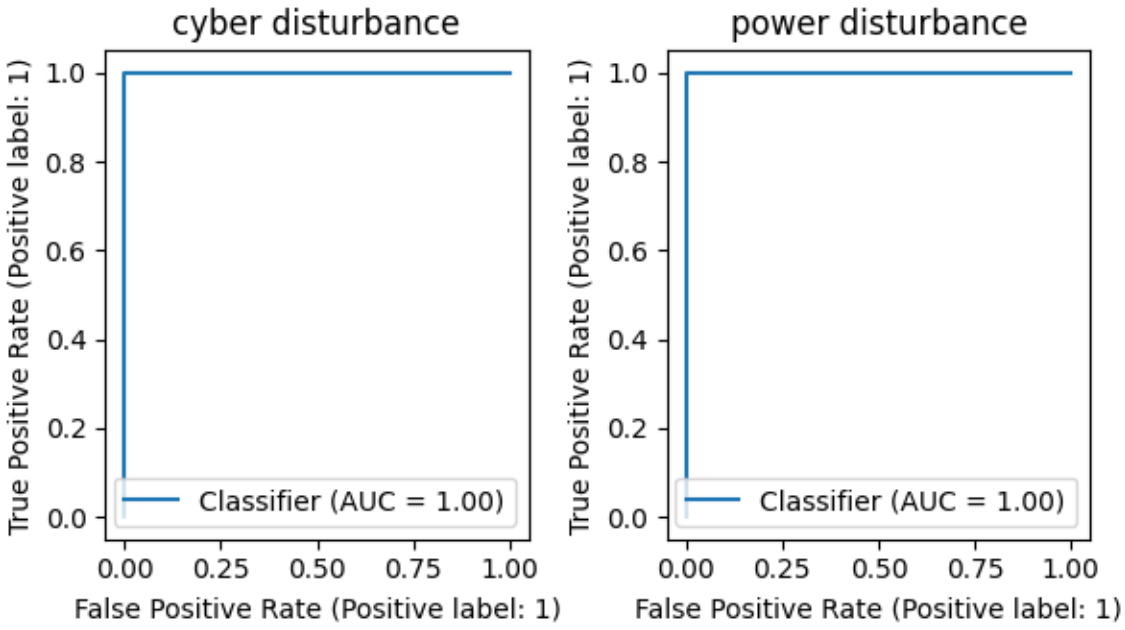


Figure A- 4

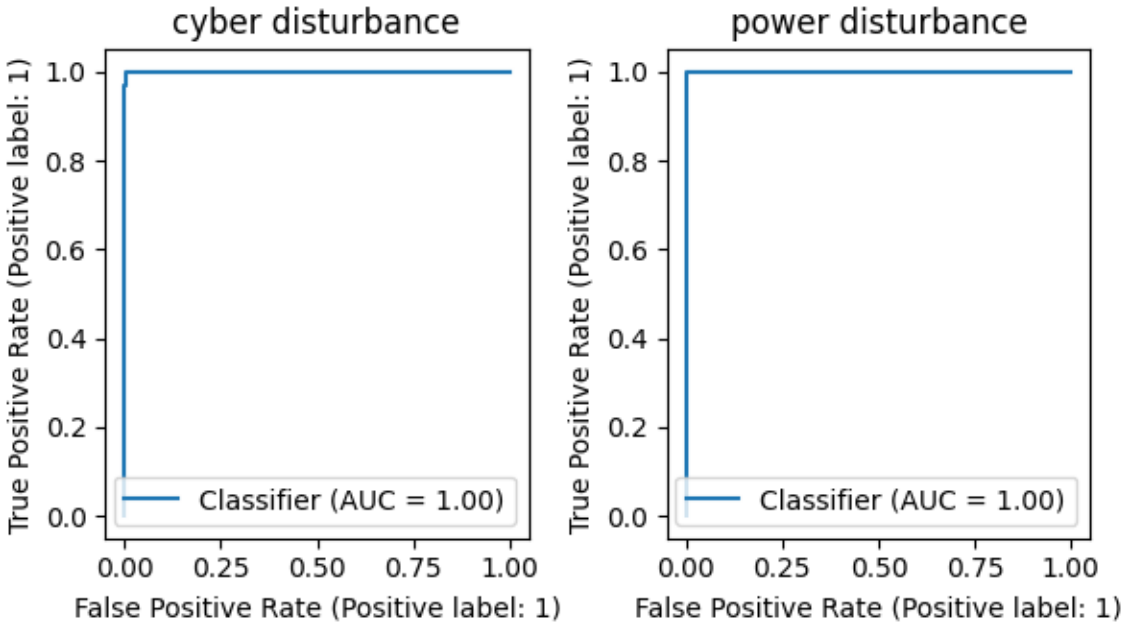


Figure A- 5

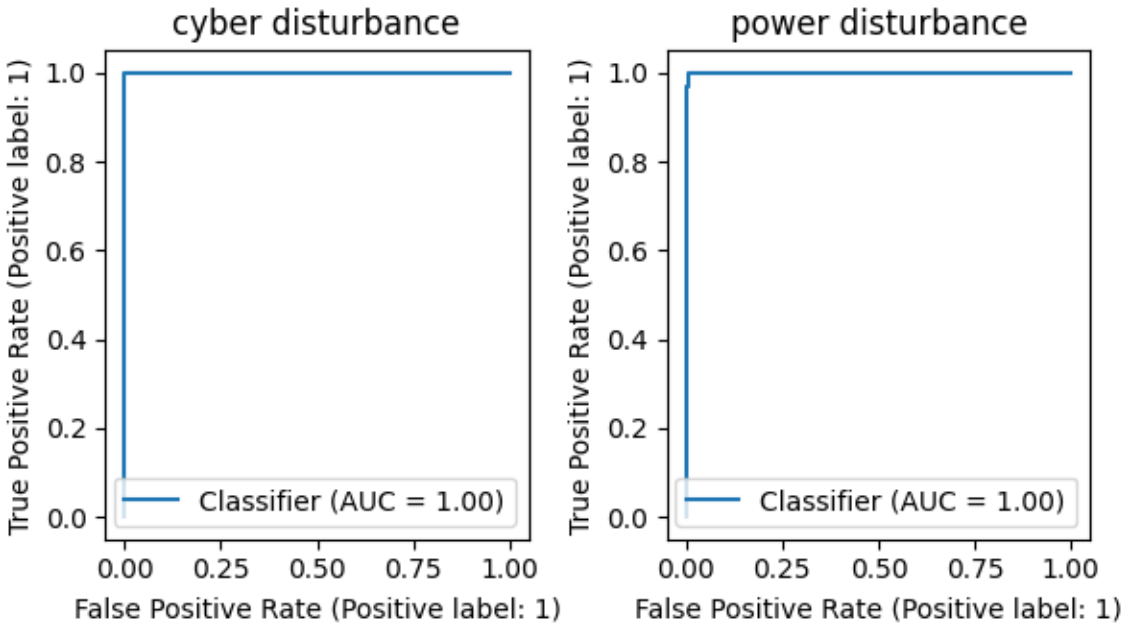

cyber disturbance
power disturbance
True Positive Rate (Positive label: 1)
False Positive Rate (Positive label: 1)
Classifier (AUC = 1.00)
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00

Figure A- 6

Figure A-7, Figure A-8, and Figure A-9-- probability histograms (Figure 5) for replicates 2-4:

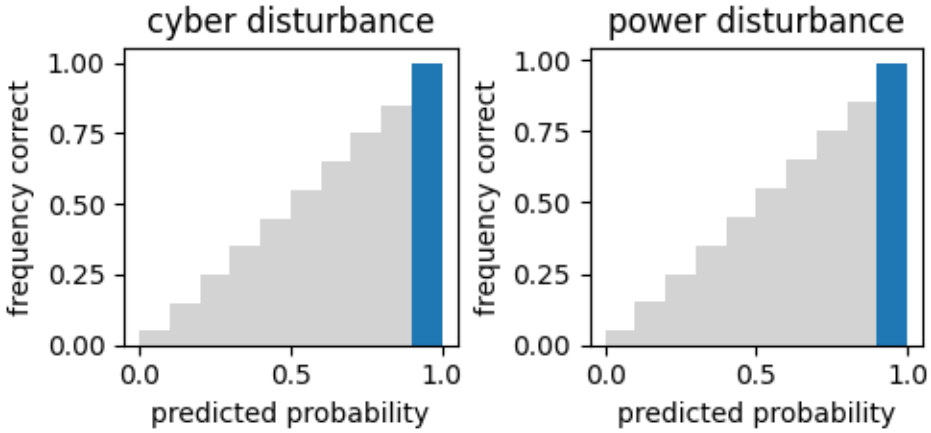


Figure A- 7

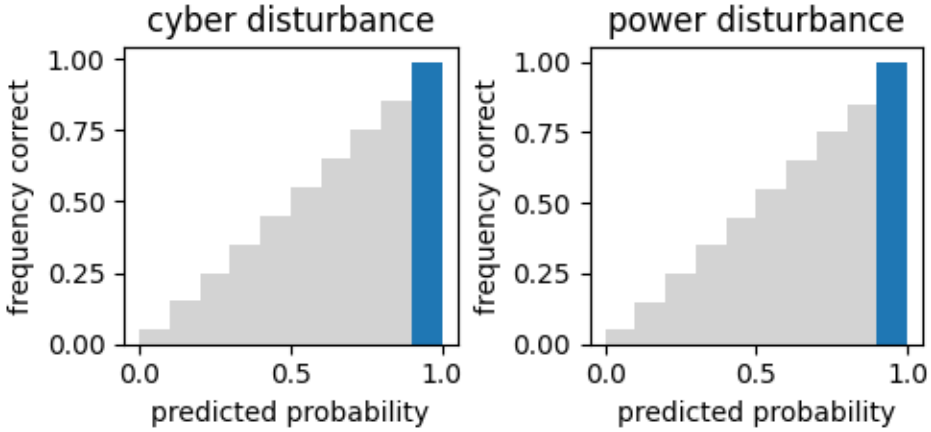


Figure A- 8

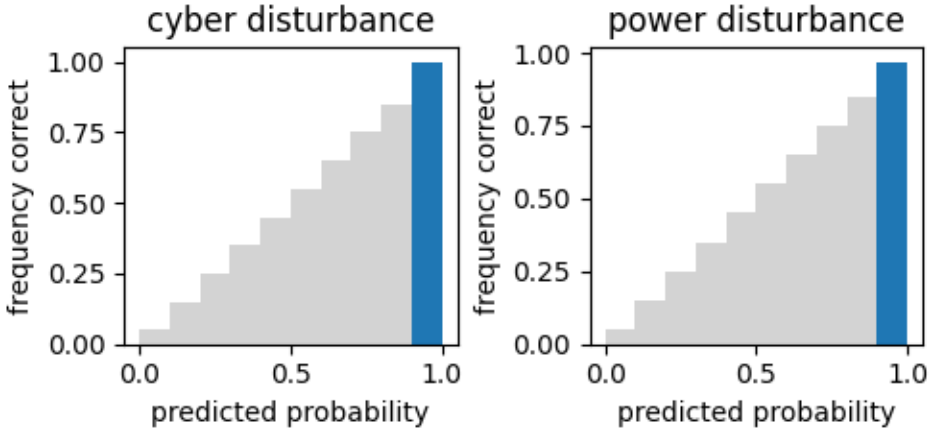


Figure A- 9

APPENDIX B. ADAPTING AUTORAS TO HARMONIE-SPS

The work on the automation of the remedial action schemes was tested on the WSCC 9-bus system, which is a transmission network approximation of the Western Electricity Coordination Council (WSCC). This 9-bus system consists of three generators, three loads, and one transformer at each of the generator units for a total of three. It is also worth noting that the base voltages of this system are at the following levels: 13.8 kV, 16.5 kV, 18 kV, and 230 kV [Ref. B-1]. An image of the system simulated on PowerWorld is shown in Figure B-1.

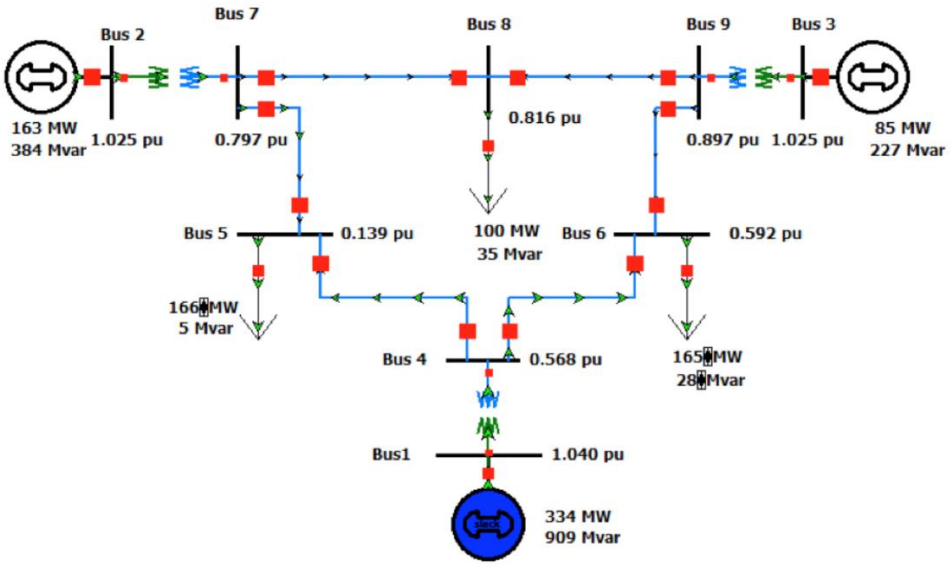


Figure B- 1: WSCC 9-Bus System

With that being said, an overview of the Automated Remedial Action Schemes (autoRAS) concept and implementation will be discussed. The main goal of implementing the automation of remedial action schemes (RAS) in this project is to ease the method of determining corrective actions in the hopes of reducing violations that result in line overloading [Ref. B-2]. To do so, the initial step was to create multiple contingency scenarios using PowerWorld Simulator, which will be discussed in more details later on. Once the contingencies were created, only those that have resulted in line overloading violations were considered. Afterward, Support Vector Machine (SVM) clustering techniques were implemented to group the contingencies in separate clusters. Each cluster includes contingency violations that can be solved with one corrective action. This implies that, in power systems, it is possible to achieve the same violation in different methods. For example, dropping the generation in either one of the generators at Bus 2 or Bus 3 could result in line overloading of the line connection between buses 8 and 9. As such, it is important to create a list of all the possible contingencies to ensure a thorough analysis of the system. However, since that results in large amounts of data, the SVM clustering techniques help to reduce that data into clusters that can then be used as inputs to the autoRAS algorithm, which can provide the mitigative corrective actions to alleviate the line overloading.

In more details, the autoRAS algorithm can be divided into two main parts, which are the Conditions and Corrective Actions. In Conditions, all the required input data is collected to serve as input to the second part which yields the corrective actions (suggested mitigations).

The first step in this algorithm is generating the contingency list and perform any data preprocessing required. To create a list of the contingencies, both PowerWorld Simulator and Easy SimAuto (ESA) were utilized. Easy SimAuto (ESA) is a python package that enables users to easily interact with Power World Simulator to collect data and run simulations [Ref. B-3]. A yearly load profile with a one-hour time step was used. The yearly load profile included hourly values of real power and reactive power in MWs and MVARs, respectively, for all three loads in the WSCC 9-bus system. PowerWorld was used to create a contingency list that includes both n-1 and n-2 contingencies. The n-1 contingencies included the following:

1. Loss of one generator
2. Loss of one transformer
3. Opening of a load
4. Opening of a line
5. Opening of a bus

The n-2 contingencies made up all possible combinations of the n-1 contingencies list above, which was done using the Contingency Analysis tool on PowerWorld Simulator. The contingencies created follow the North American Electric Reliability Corporation (NERC) standard of TPL-001-5, which is named "Transmission System Planning Performance Requirements" [Ref. B-4]. Following this standard, the contingencies that were considered as classified as P1 (single contingency), P2 (single contingency), and P6 (multiple contingencies). The P1 and P6 contingencies could be a 3-phase or a single line-to-ground (SLG) fault, while the P2 contingency can only be a SLG fault. The tables below provide a more detailed description of the different categories of contingencies under this standard [B-4].

Category	Initial Condition	Event [1]	Fault Type [2]	BES Level [3]	Interruption of Firm Transmission Service Allowed [4]	Non-Consequential Load Loss Allowed
P0 No Contingency	Normal System	None	N/A	EHV, HV	No	No
P1 Single Contingency	Normal System	Loss of one of the following: 1. Generator 2. Transmission Circuit 3. Transformer[5] 4. Shunt Device[6]	3Ø	EHV, HV	No[9]	No[12]
		5. Single Pole of a DC line	SLG			
P2 Single Contingency	Normal System	1. Opening of a line section w/o a fault [7]	N/A	EHV, HV	No[9]	No[12]
		2. Bus Section Fault	SLG	EHV	No[9]	No
				HV	Yes	Yes
		3. Internal Breaker Fault[8] (non-Bus-tie Breaker)	SLG	EHV	No[9]	No
				HV	Yes	Yes
		4. Internal Breaker Fault (Bus-tie Breaker)[8]	SLG	EHV, HV	Yes	Yes

Figure B- 2: NERC Contingency Categories (Part 1)

Category	Initial Condition	Event [1]	Fault Type [2]	BES Level [3]	Interruption of Firm Transmission Service Allowed [4]	Non-Consequential Load Loss Allowed
P5 Multiple Contingency *(Fault plus non-redundant component of a Protection System failure to operate)*	Normal System	Delayed Fault Clearing due to the failure of a non-redundant component of a Protection System[13] protecting the Faulted element to operate as designed, for one of the following: 1. Generator 2. Transmission Circuit 3. Transformer[5] 4. Shunt Device[6] 5. Bus Section	SLG	EHV	No[9]	No
				HV	Yes	Yes
P6 Multiple Contingency *(Two overlapping singles)*	Loss of one of the following followed by System adjustments.[9] 1. Transmission Circuit 2. Transformer [5] 3. Shunt Device[6] 4. Single pole of a DC line	Loss of one of the following: 1. Transmission Circuit 2. Transformer[5] 3. Shunt Device[6]	3Ø	EHV, HV	Yes	Yes
		4. Single pole of a DC line	SLG	EHV, HV	Yes	Yes

Figure B- 3: NERC Contingency Categories (Part 2)

Once the initial contingency list was created through PowerWorld, the contingencies then were evaluated to study their effect on the system. Utilizing the load profile, all the contingencies created in the previous step were run at every time step in the load profile. This means that for every different value of load (system operational snapshot), the contingencies were run to determine if they would cause line overloading violations. This process resulted in a large number of contingencies, where some of which did not cause any system violations. The remaining contingencies that resulted in violations were combined in .csv files.

Once the contingency list was finalized, separate scenarios were created which correspond to the different normal system operational values at the different load values from the yearly load profile. A larger file was then created that has all the operational values of the system for all the different scenarios. These values include power flow on the different branches in the system, the per unit voltages at the buses and the voltage angles, and the real and reactive power injections at the buses. Once the scenario normal operational values and the contingency list were finalized, the violations were determined by using ESA to help run the contingency analysis. The violations that were studied were on the line MVA values to help determine line overloading. The collected data includes the value of the violation and the branch. Once that was determined, distance-based SVM clustering was implemented to cluster the contingencies that led to similar violations that would require the same corrective action.

With that being said, the algorithm then moves to the corrective actions part. The suggested mitigations include dropping generation, load shedding, or branch switching to help alleviate the line overloading. It is important to note that due to the SVM clustering, this algorithm was only successful in running and analyzing n-1 contingencies.

When the corrective actions were suggested, they were then tested on PowerWorld Simulator by manually implementing the suggested corrective actions. When that was done, it was observed that the corrective actions changed the MW values of the generation, but the reactive power values were not included, causing the slack bus to make up for that reactive power. As such, it was concluded that not including the reactive power in the autoRAS algorithm when suggesting the corrective actions still leads to some line overloading even when the mitigations are implemented. With that being said, changes were made to the original autoRAS algorithm to help take into account the reactive power when providing corrective actions. The reactive power is now included in the autoRAS algorithm, which resulted in better mitigative actions. Regardless, line overloading is still apparent in certain cases even after the mitigative actions were implemented. With that being said, suggested future work should include studying the algorithm more and possibly creating an algorithm that generates a second round of corrective actions that can alleviate the line overloading as much as possible.

Last but certainly not least, it is important to study the input data that goes into the SVM/autoRAS code pipeline to be able to incorporate it with the larger ML framework in this project. Since the autoRAS algorithm has a dependency on PowerWorld Simulator and ESA, it would be run in parallel to the developed ML framework. This limitation exists due to the fact that the PowerWorld Simulator can only operate on a Windows-based machine. While taking that into account, it was still concluded that it would be best to make sure the input data files that are trained on the autoRAS algorithm and the ML framework be the same. As such, the contingency lists and input files were adjusted accordingly.

To achieve this, it was first important to look into the placement of the phasor measurement unit (PMU) devices in the system that collect the data for the ML framework. It is important to note that when simulating a system on PowerWorld, the system observability is essentially 100%, as you can collect all the data that is needed in all the locations. However, when it comes to PMU collected data, a limitation arises depending on where the PMU is located. As such, the PMU placements for the WSCC 9-bus system are as follows:

1. PMU1: located at Bus 7
2. PMU2: located at Bus 8 Load
3. PMU3: located at Bus 9
4. PMU4: located at Bus 5 Load
5. PMU5: located at Bus 5
6. PMU6: located at Bus 4
7. PMU7: located at Bus 6
8. PMU8: located at Bus 6 Load

This meant that the data to be collected from PowerWorld to only be collected from the locations mentioned above to match what the PMU devices reported. One other important aspect to mention is that PMU devices collected more values of data than what was originally collected in the initial contingency list. The additional collected values include:

1. system frequency

2. voltage magnitude and angle
3. current magnitude and angle

The system frequency and the voltage magnitude and angle values were directly collected from PowerWorld Simulator. However, for the current magnitude and angle, calculations were performed from the injected power values and voltage values at each of the buses. After all the adjustments were made, the input files to the autoRAS algorithm match the input to the ML framework. Once that was achieved, data reformatting and preprocessing was performed to make the input compatible with the autoRAS algorithm, and analysis was run in a similar manner as described above to obtain the corrective actions from the algorithm. A future work suggestion can include the comparison of the accuracy and precision of the autoRAS algorithm when inputting all the data collected from PowerWorld against the partial system data collected from the PMU devices.

References:

- Ref. B-1. "WSCC 9-Bus System," *Texas A&M University College of Engineering*. [Online]. Available: https://electricgrids.engr.tamu.edu/electric-grid-test-cases/wscc-9-bus-system/.

- Ref. B-2. H. Li, K. Shetye, T. Overbye, K. Davis, and S. Hossain-Mckenzie, "Towards the automation of remedial action schemes design," *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2021.

- Ref. B-3. B. Thayer, Z. Mao, Y. Liu, K. Davis, and T. Overbye, "Easy SimAuto (ESA): A python package that simplifies interacting with PowerWorld Simulator," *Journal of Open Source Software*, vol. 5, no. 50, p. 2289, 2020.

- Ref. B-4. "TPL-001-5: Transmission System Planning Performance Requirements," *The North American Electric Reliability Corporation (NERC)*. [Online]. Available: https://nerc.com/pa/Stand/Reliability%20Standards/TPL-001-5.pdf.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Derek Hart	5621	derhart@sandia.gov
Craig Lawton	8141	crlawto@sandia.gov
LDRD office	1910	ldrd@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.